

# РУКОВОДСТВО ПО ОЦЕНКЕ КАЧЕСТВА ПОИСКА



Google

# СОДЕРЖАНИЕ

Версия 1.0 2 ноября 2012

## ПРЕДИСЛОВИЕ

<b>Глава I: ПРИНЦИПЫ ОЦЕНКИ.....</b>	<b>4</b>
<b>1. Добро пожаловать в программу по оценке качества поиска!.....</b>	<b>4</b>
1.1 Обзор оценки URL.....	4
1.2 Важные термины и идеи.....	4
1.3 Цель оценки качества поиска.....	6
1.4 Аессоры должны представлять пользователя.....	6
1.5 Информация о безопасности в интернете.....	6
<b>2. Когда непонятен запрос.....</b>	<b>7</b>
2.1 Когда непонятно намерение пользователя.....	7
2.2 Язык и местоположение задачи.....	8
2.3 Многозначные запросы.....	8
2.4 Классификация запросов: транзакционные, информационные и навигационные (или Do-Know-Go).....	9
<b>3. Язык целевой страницы.....</b>	<b>14</b>
<b>4. Оценка по шкале релевантности.....</b>	<b>14</b>
4.1 Витальная.....	15
4.2 Полезная.....	16
4.3 Релевантная.....	16
4.4 Относительно релевантная.....	16
4.5 Нерелевантная/Неполезная.....	16
4.6 Нерейтингуемая.....	17
<b>5. От намерения пользователя до выставления оценки.....</b>	<b>20</b>
5.1 Намерение пользователя и полезность страницы.....	20
5.2 О важности местоположения.....	21
5.3 О важности языка.....	21
5.4 Многозначные запросы.....	22
5.5 Четкость запросов и целевых страниц.....	24
5.6 Распространенные ошибки при оценке страниц.....	25
<b>6. Метки.....</b>	<b>29</b>
6.1 Метка Спам.....	30
6.2 Метка Порно.....	30
6.3 Метка Вредоносная страница.....	31
6.4 Совместимость оценок и меток.....	31



Понравилась книга? Расскажи друзьям!

# СОДЕРЖАНИЕ

<b>Глава II: ОЦЕНКА URL С УЧЕТОМ РЕГИОНА ЗАПРОСА.....</b>	<b>32</b>
1. Регион запроса.....	32
2. Оценка страницы под геозависимый запрос.....	34
3. Принцип оценки при наличии региона запроса.....	35
4. Когда важен регион запроса?.....	35
<b>Глава III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ .....</b>	<b>38</b>
1. Что такое поисковый спам.....	38
1.1 Взаимосвязь оценок и спама.....	38
1.2 Зачем создают спам-страницы.....	38
1.3 Когда проверять спам.....	39
2. Поиск технических сигналов .....	39
2.1 Скрытый текст и скрытые ссылки.....	39
2.2 Перенасыщение ключевыми словами.....	41
2.3 Скрытые редиректы.....	42
2.4 Клоакинг.....	43
3. Полезные страницы vs спам-страницы.....	43
3.1 Страницы с копированным текстом и PPC-рекламой.....	44
3.2 Фейковые страницы поиска с PPC-рекламой.....	46
3.3 Фейковые блоги с PPC-рекламой.....	46
3.4 Фейковые доски объявлений с PPC-рекламой.....	46
3.5 Копированный контент, который НЕ является спамом.....	47
4. Коммерческие намерения.....	47
4.1 Неполные аффилиаты.....	47
4.2 Страницы только с PPC-рекламой.....	48
4.3 Припаркованные/просроченные домены.....	48
4.4 Страницы с бесполезным содержанием и PPC-рекламой.....	49
5. Фишинговые сайты.....	50
6. Спам и стадия решения.....	50
7. Заключение.....	51

В своей деятельности Google полагается на армию асессоров, работающих во многих странах и на многих языках мира. Именно они помогают определять качество поисковой выдачи, ранжирования и целевых страниц. Асессоры выполняют самые разные «оценочные действия», призванные предоставить нам информацию о качестве результатов на разные типы запросов. Собранная ими информация позволяет команде Google взглянуть на качество поиска во временном разрезе, а также оценить эффект от изменений, ранее внесенных в поисковый алгоритм Google. Оценки асессоров не оказывают прямого влияния на ранжирование результатов в выдаче Google. Асессор может присвоить тому или иному URL определенный балл, но эта оценка напрямую не повлияет на позицию сайта в поисковой выдаче. В свою очередь, оценка асессора принимается во внимание при анализе качества поиска и решении внести изменения в поисковый алгоритм.

Данный документ — краткое описание наших принципов оценки качества поиска. Обратите внимание на то, что это не полная версия руководства, которое асессоры используют

в своей работе; это лишь краткое изложение важных моментов. Версия асессоров включает в себя инструкции по использованию рейтинговых интерфейсов, дополнительных оценочных примеров и т.д. Данное руководство представляет собой описание оценочных действий для асессоров и фокусируется на ключевом моменте — оценке URL. При оценке URL асессору показывается запрос согласно его местоположению (страна+язык) и URL, который может быть выдан поисковиком на данный запрос. Асессор оценивает качество результата для данного запроса по специальной шкале, представленной в документе. Звучит просто, не правда ли? Однако, как вы сможете заметить, существует множество случаев, где приходится подумать, прежде чем вынести оценку. Данный документ помогает асессорам правильно взглянуть на некоторые из таких ситуаций.

Наши принципы оценки качества поиска постоянно меняются, потому что мы изучаем поиск и развиваем его. Эту версию документа мы разработали специально для тех людей, которые хотят больше узнать о том, как Google понимает релевантность и качество поисковых результатов.

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

## 1. ДОБРО ПОЖАЛОВАТЬ В ПРОГРАММУ ПО ОЦЕНКЕ КАЧЕСТВА ПОИСКА!

Вы как эксперт по качеству поиска, асессор, будете работать с различными типами оценочных проектов. Однако нижеизложенные принципы распространяются только на один такой проект — оценку URL.

Пожалуйста, внимательно изучите все принципы. Представленные здесь идеи понадобятся вам для других типов оценочных проектов. Понимание того, как оценивать URL, поможет вам стать хорошим асессором.

### 1.1 ОБЗОР ОЦЕНКИ URL

Для каждого задания по оценке URL вам будут даны запрос и ссылка. Вам будет нужно:

- Изучить данный запрос.
- Перейти по ссылке на целевую страницу.
- Вынести оценку по шкале релевантности.

### 1.2 ВАЖНЫЕ ТЕРМИНЫ И ИДЕИ

**Поисковая система (search engine):** Поисковая система — это сайт, позволяющий пользователям осуществлять поиск в сети посредством введения слов или символов в поисковую строку.

**Запрос (query):** Запрос — это набор слов, чисел и/или символов, которые пользователь вводит в поисковую строку поисковика. Иногда для обозначения такого набора слов, чисел и/или символов мы используем термин «условия запроса». Также применяется термин «ключевые слова». В данном документе запрос будет выделяться квадратными скобками, например, [цифровые камеры].

**Намерение пользователя (user intent):** Когда пользователь вводит запрос, он хочет найти определенную информацию или купить что-то онлайн. Именно эту цель мы называем намерением пользователя.

**Язык и местоположение задачи (task language and task location):** Для всех запросов определяется язык задачи и местоположение, в данном документе это будет обозначаться следующим образом: [digital cameras], Spanish (ES). Данный формат означает, что запрос digital cameras был введен в строку поиска испаноговорящим пользователем в Испании. Местоположение обозначается двухбуквенным кодом страны: для Испании это ES. Если бы запрос был введен испаноговорящим пользователем в Мексике, мы бы имели следующее обозначение: [digital cameras], Spanish (MX).

Актуальные коды стран вы найдете здесь [http://www.iso.org/iso/country\\_codes/iso\\_3166\\_code\\_lists/country\\_names\\_and\\_code\\_elements.htm](http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm).

**Домашняя страница (homepage):** Под термином «домашняя страница» мы понимаем главную страницу сайта. Это первая страница, которую видит пользователь после загрузки сайта. URL домашней страницы сайта обычно заканчивается на .com, .edu, .org, .gov или на двухбуквенный код страны: .jp, .mx, .ru и т.п. Например, <http://www.apple.com/> — это домашняя страница сайта компании Apple. Отметим, что в некоторых странах под «домашней страницей» понимают весь сайт. В данном документе этот термин используется только для обозначения главной страницы сайта.

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

**Внутренняя страница (subpage):** Страница сайта, не являющаяся домашней. Например, <http://www.apple.com/iphone/> — это внутренняя страница сайта Apple.

**Веб-страница (webpage):** Любая страница сайта.

**URL:** Адрес веб-страницы, которую оценивают асессоры. Например, <http://www.microsoft.com>. Важно обращать внимание на внешний вид URL/ссылки, но помните: оценивается именно целевая страница.

**Целевая страница (landing page):** Именно эту веб-страницу оценивают асессоры. Это та страница, которую видит пользователь после перехода по URL из поиска. В данном руководстве вы найдете правила, по которым оценивается целевая страница. На многих целевых страницах вы столкнетесь с рекламой и платными ссылками, однако вам будет нужно оценивать только контент целевой страницы. Реклама и ссылки (даже если они тематические) не должны повлиять на вашу оценку.

**Тематика (topic):** Это тема/предмет запроса. Когда пользователи ищут ответ на свой запрос, они хотят найти страницы, относящиеся к теме запроса.

**Полезность (utility):** Полезность целевой страницы — это показатель, который говорит о том, насколько страница отвечает намерениям пользователя. И чем выше этот показатель, тем больше ее польза для юзера.

**Полезность — наиболее важный аспект в качестве поиска, и поэтому вы должны очень серьезно подойти к ее оценке.**

**Шкала релевантности (rating scale)** будет подробно рассмотрена в разделе 4 данной главы. Здесь же краткий обзор. Для каждого задания вам будет необходимо указать одну из следующих оценок по шкале релевантности:

ОЦЕНКА РЕЛЕВАНТНОСТИ	ОПИСАНИЕ
Витальная	Особая оценочная категория. Подробнее в главе 4.1.
Полезная	Страница, которая является ценой для большинства пользователей.
Релевантная	Страница, которая является полезной для многих или некоторых людей.
Относительно релевантная	Страница, которая является не очень ценной для большинства пользователей, но в некоторой степени относится к тематике запроса. Небольшое количество юзеров может найти ее полезной для себя.
Нерелевантная/Неполезная	Страница, которая полезна слишком малому количеству людей либо никому.
Нерейтингуемая	Страница, которую невозможно оценить. Подробнее в главе 4.6.

Помимо оценок, страницам нужно будет присваивать следующие метки: **Не Спам, Возможно, спам, Спам, Порно и Вредоносная страница.**

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 1.3 ЦЕЛЬ ОЦЕНКИ КАЧЕСТВА ПОИСКА

Ваши оценки будут использованы для определения качества поиска Google по всему миру. Хорошие поисковые системы предоставляют такие результаты поиска, которые ценные для пользователей с учетом их языка и местонахождения. Обратите внимание, что ваши оценки не оказывают прямого воздействия на ранжирование результатов в поиске/ на алгоритмы ранжирования Google.

### 1.4 АСЕССОРЫ ДОЛЖНЫ ПРЕДСТАВЛЯТЬ ПОЛЬЗОВАТЕЛЯ

Очень важно, чтобы вы представляли пользователя поисковика. Пользователь — это тот, кто живет в заданном географическом регионе и читает на заданном языке, тот, кто ввел запрос в поисковую строку Google.

Вы должны быть хорошо знакомы с языком и местоположением задачи, чтобы понимать поисковые практики пользователя в заданном вам месте. Если вы не обладаете соответствующими знаниями, обязательно проинформируйте об этом вашего работодателя.

### 1.5 ИНФОРМАЦИЯ О БЕЗОПАСНОСТИ В ИНТЕРНЕТЕ

В ходе своей работы вы будете посещать много различных веб-страниц. Некоторые из них могут оказаться небезопасными для вашего компьютера. Во избежание этого не скачивайте EXE-файлы и различные приложения или другие потенциально опасные программы, не переходите по ссылкам, в которых вы не уверены. Мы настоятельно рекомендуем установить на компьютер антивирусные и антишпионские программы. Данные программы безопасности должны регулярно обновляться. В сети вы сможете найти большое количество как платных, так и бесплатных вариантов антивирусов.

Ниже приведены ссылки на статьи про антивирусные программы:

[http://en.wikipedia.org/wiki/Antivirus\\_software](http://en.wikipedia.org/wiki/Antivirus_software)  
<http://en.wikipedia.org/wiki/Spyware>

Пропустите оценочное задание, если вы видите в нём неизвестные или подозрительные форматы файлов. Ниже перечислены форматы файлов надежные при условии, что антивирус установлен:

- **.txt (текстовый файл),**
- **.ppt или .pptx (Microsoft PowerPoint),**
- **.doc или .docx (Microsoft Word),**
- **.xls или .xlsx (Microsoft Excel),**
- **.pdf (PDF).**

Если вам встретилась страница, которую поиск/браузер или антивирус рекомендуют не посещать, вы вправе это сделать и отметить данную страницу следующим образом — **Нерейтингаемая: Не загрузилась**. Подробней о данной оценке в разделе 4.6.1 первой главы.

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Также вам могут встретиться страницы, которые потребуют RealPlayer или Adobe Flash Player. Вы можете загрузить их по ссылкам:

<http://www.real.com/>

[http://www.adobe.com/shockwave/download/download.cgi?P1\\_Prod\\_Version=ShockwaveFlash](http://www.adobe.com/shockwave/download/download.cgi?P1_Prod_Version=ShockwaveFlash)

А вот примеры страниц, которые требуют установки flash-плеера, <http://www.ferrariworld.com> и <http://www.atraircraft.com>.

## 2. КОГДА НЕПОНЯТЕН ЗАПРОС

Прежде чем оценивать страницу, вам необходимо понять запрос. Для этой цели вы можете использовать онлайн-словари и энциклопедии. А если понадобится, то исследовать данный вопрос, чтобы наверняка разобраться в запросе пользователя.

**Важно:** Если вы применяете поисковую систему для изучения запроса пользователя, не полагайтесь только на позиции ссылок в выдаче. Запрос может оказаться многозначным, а в первых результатах может быть не то значение слова, которое вам нужно. Не оценивайте страницу высоко только потому, что она появилась в ТОПе поисковой выдачи.

Полезные ссылки:

Онлайн-энциклопедии

[http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page): английская версия «Википедии».

<http://www.wikipedia.org/>: здесь можно выбрать нужный язык.

Онлайн-переводчики

<http://babelfish.yahoo.com/>

<http://www.wordreference.com/>

<http://translate.google.com/>

### 2.1 КОГДА НЕПОНЯТНО НАМЕРЕНИЕ ПОЛЬЗОВАТЕЛЯ

Помимо сути запроса, вы должны понимать и намерение пользователя. Какое намерение было у человека, когда он вбивал тот или иной запрос? Вы должны понимать это, чтобы правильно оценить целевую страницу.

Рассмотрим запрос [tetris], English (US). Большинство пользователей, вводящих такой запрос, знают, что «Тетрис» — это компьютерная игра. Поэтому пользовательским намерением, скорее всего, является желание поиграть в нее онлайн.

Вот еще примеры запросов и намерений:

Запрос	Возможное намерение пользователя
[Fedex], English (US)	Отследить посылку или найти местоположение почтового офиса.
[calendar], English (US)	Найти и распечатать календарь на текущий месяц/год. Найти календарь праздников. Найти онлайн-календарь.
[ebay], English (US)	Купить или продать что-то на eBay. Изучить главную страницу интернет-аукциона.



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 2.2 ЯЗЫК И МЕСТОПОЛОЖЕНИЕ ЗАДАЧИ

У всех запросов есть язык задачи и ее местоположение. Помня об этом, вы быстрее разберетесь с запросом и намерением пользователя. Люди из разных концов света могут вкладывать разные понятия в один и тот же запрос.

Запрос	Значение запроса согласно местоположению	Возможное намерение пользователя данного местоположения
[football], English (US)	В американский футбол играют коричневым овальным мячом.	Найти счет последних игр, расписание игр, фотографии, данные о команде и т.д. для американского футбола в США.
[football], English (UK)	В эту игру играют круглым мячом. Американцы называют ее soccer.	Найти счет последних игр, расписание игр, фотографии, данные о команде и т.д. для футбола в Великобритании или по всему миру.

### 2.3 МНОГОЗНАЧНЫЕ ЗАПРОСЫ

Существует много запросов, которые имеют больше одного значения. Например, под англоязычным запросом [apple] можно подразумевать как фрукт, так и название компании. Возможные значения запроса мы называем **интерпретациями запроса**.

**Доминирующая интерпретация:** Это наиболее частая интерпретация запроса, которую применяют пользователи. Например, большинство пользователей, ищущих [windows], имеют в виду операционную систему Microsoft, а не окна.

**Общая интерпретация:** В некоторых случаях у запроса не бывает доминирующей интерпретации. Например, под англоязычным запросом [mercury] могут иметь в виду и автомобильный бренд, и планету, и химический элемент (Hg). В данном случае ни один из вариантов не доминирует, т.е. подобные запросы имеют общую интерпретацию.

**Второстепенная интерпретация:** Иногда вы будете сталкиваться с менее распространенными интерпретациями запроса. Это значения, которые используются очень небольшим числом пользователей. Именно такие значения мы называем второстепенными интерпретациями. Приведем тот же пример: англоязычный запрос [mercury], (США). Среди его возможных значений существуют такие, о которых знает не каждый американский пользователь, например, страховая компания Mercury Marine Insurance и газета San Jose Mercury News. Это и есть второстепенные интерпретации.

При оценке страницы под второстепенную интерпретацию запроса вы будете использовать не высокие оценки из шкалы релевантности. В разделе 5.4 данной главы мы рассмотрим такие случаи более детально.

### 2.4 КЛАССИФИКАЦИЯ ЗАПРОСОВ: ТРАНЗАКЦИОННЫЕ, ИНФОРМАЦИОННЫЕ И НАВИГАЦИОННЫЕ (ИЛИ DO-KNOW-GO)

Иногда полезно классифицировать запросы по данным типам:

**Транзакционные запросы** — пользователи хотят совершить некоторое действие. Например, скачать программу, сыграть в онлайн-игру, послать цветы, найти видео и т.д. Это запросы действия или do-запросы.

**Информационные запросы** — пользователи ищут нужную им информацию. Это так называемые know-запросы.

**Навигационные запросы** — пользователи ищут определенную страницу или сайт. Это go-запросы, т.к. пользователь собирается перейти на определенную страницу.

Do-know-go — легкий способ запомнить вышепредставленную классификацию запросов, которая поможет вам в оценке страниц. Только учитывайте, что многие запросы могут относиться сразу к нескольким типам.

#### 2.4.1 ТРАНЗАКЦИОННЫЕ ЗАПРОСЫ — DO

Цель транзакционного запроса — совершить определенное действие в сети: купить, загрузить, получить и т.д. Пользователь хочет **сделать** что-то. Вот примеры возможных действий:

- Купить продукт;
- Загрузить программу бесплатно или за деньги;
- Оплатить счет онлайн;
- Поиграть онлайн;
- Распечатать календарь;
- Послать цветы;
- Посмотреть видеоклип;
- Скопировать картинку;
- Провести онлайн-опрос;
- Посмотреть интересные сайты.

Полезными для транзакционного запроса являются страницы, позволяющие пользователю совершить нужное действие и достичь цели.

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

ЗАПРОС	ВОЗМОЖНОЕ НАМЕРЕНИЕ ПОЛЬЗОВАТЕЛЯ	ОПИСАНИЕ ЦЕЛЕВОЙ СТРАНИЦЫ
[geography quiz], English (US)	Пройти онлайн-тест по географии	Страница с онлайн-тестом по географии
[Beatles poster], English (US)	Найти картинку с плакатом «Битлз» или купить плакат «Битлз»	Страница, на которой можно посмотреть или купить плакат «Битлз»
[download adobe reader], English (US)	Загрузить программу	Страница с бесплатной загрузкой на официальном сайте Adobe
[fairy tale coloring pages], English (US)	Распечатать раскраски	Надежная страница с раскрасками, которые можно распечатать
[online personality test], English (US)	Пройти личностный тест онлайн	Страница с личностным тестом
[what is my bmi?], English (US)	Рассчитать индекс массы тела	Страница с работающим калькулятором расчета
[good cop baby cop], English (US)	Посмотреть видео "Плохой хороший полицейский"	Страница, на которой можно посмотреть видео
[cute kitten pics], English (US)	Посмотреть фотографии котят	Страница с фотографиями котят
[Citizen Kane DVD], English (US)	Купить данный DVD-диск	Страницы, на которых можно купить данный DVD-диск
[flowers], English (US)	Заказать цветы онлайн	Страница, на которой можно сделать онлайн-заказ цветов
[play sudoku], English (US)	Играть в судоку онлайн	Страница, где можно сыграть в судоку
[calculate running pace], English (US)	Рассчитать скорость бега онлайн	Страница с работающим онлайн-калькулятором
[bubble spinner 2], English (US)	Играть онлайн или загрузить игру	Страница, на которой можно играть онлайн и/или загрузить игру
[Spanish English dictionary], English (US)	Перевести слово с испанского на английский или наоборот	Страница, на которой можно осуществить перевод

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 2.4.2 ИНФОРМАЦИОННЫЕ ЗАПРОСЫ — KNOW

Информационный запрос требует информации по заданной теме. Пользователю нужно что-то узнать, его цель — найти информацию.

Полезные страницы содержат качественные, авторитетные и исчерпывающие данные о запросе.

Запрос	Возможное намерение пользователя	Описание целевой страницы
[Switzerland], English (US)	Найти информацию для планирования поездки. Или найти информацию о географии, языке, экономике Швейцарии.	Страница о Швейцарии на сайте известного путеводителя.
[cryptology use in WWII], English (US)	Найти информацию о том, как использовалась криптология во время Второй мировой войны.	Статья о криптологии во времена Второй мировой войны с сайта Музея ВВС США.
[how to remove candle wax from carpet], English (US)	Найти информацию о том, как удалить с ковра воск.	Страница на данную тему с сайта известного журнала.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 2.4.3 НАВИГАЦИОННЫЕ ЗАПРОСЫ — GO

Цель навигационного запроса — найти конкретную страницу в сети. При поиске пользователи думают о конкретной странице или конкретном сайте. Такая страница является целью запроса.

Наиболее полезной страницей для навигационного запроса является страница компании или ресурс, который искал пользователь.

Запрос	Возможное намерение пользователя	URL целевой страницы	Описание целевой страницы
[ibm], English (US)	Перейти на домашнюю страницу IBM	<a href="http://www.ibm.com/">http://www.ibm.com/</a>	Главная страница официального сайта Корпорации IBM
[youtube], English (US)	Перейти на домашнюю страницу YouTube	<a href="http://www.youtube.com/">http://www.youtube.com/</a>	Главная страница официального сайта YouTube
[ebay], Italian (IT)	Перейти на домашнюю страницу итальянской версии eBay	<a href="http://www.ebay.it/">http://www.ebay.it/</a>	Главная страница итальянской версии сайта eBay
[harvard college admissions], French (FR)	Перейти на страницу сайта Гарвардского университета со вступительными экзаменами в колледж	<a href="http://admissions.college.harvard.edu/index.html">http://admissions.college.harvard.edu/index.html</a>	Соответствующая страница на сайте Гарвардского университета
[best buy store locator], English (US)	Перейти на соответствующую страницу сайта магазина	<a href="http://www.bestbuy.com/site/olspage.jsp?id=cat12090&amp;type=page">http://www.bestbuy.com/site/olspage.jsp?id=cat12090&amp;type=page</a>	Страница на официальном сайте Best Buy
[sony customer support], English (US)	Перейти на страницу клиентской поддержки сайта Sony	<a href="http://esupport.sony.com/">http://esupport.sony.com/</a>	Страница клиентской поддержки на официальном сайте Sony
[outback steakhouse menu], English (US)	Перейти на страницу меню сайта Outback	<a href="http://www.outback.com/menu/">http://www.outback.com/menu/</a>	Страница с меню на официальном сайте Outback
[canon.com digital cameras], English (US)	Перейти на страницу с цифровыми камерами сайта canon.com. Несмотря на то, что компания Canon известна своими цифровыми камерами, главная страница в данном случае не подойдет.	<a href="http://www.usa.canon.com/consumer/controller?act=ProductCatIndexAct&amp;fcategoryid=113">http://www.usa.canon.com/consumer/controller?act=ProductCatIndexAct&amp;fcategoryid=113</a>	Страница о цифровых камерах на сайте Canon
[facebook login], English (US)	Перейти на страницу с формой регистрации. Несмотря на то, что пользователь может зарегистрироваться с главной страницы FB, целевой страницей под данный запрос будет именно страница с регистрационной формой.	<a href="http://www.facebook.com/login.php">http://www.facebook.com/login.php</a>	Страница с формой регистрации сайта Facebook



Понравилась книга? Расскажи друзьям!

# ГЛАВА I:

## ПРИНЦИПЫ ОЦЕНКИ

### 2.4.4 МНОГОЗНАЧНЫЕ ЗАПРОСЫ (DO-KNOW-GO)

Некоторые запросы могут иметь более одного значения, которое подразумевает пользователь при поиске. Будьте внимательны, когда будете определять, какое именно значение запроса имеет в виду пользователь. Вот несколько примеров:

Запрос	Возможное намерение пользователя	URL полезной страницы	Описание целевой страницы
[download firefox], English (US)	Do и Go. Данный запрос может быть как транзакционным, так и навигационным. Пользователи хотят загрузить браузер Firefox (транзакционный запрос). Многие пользователи хотят загрузить браузер с официальной страницы сайта Firefox (навигационный запрос).	<a href="http://download.cnet.com/mozilla-firefox/">http://download.cnet.com/mozilla-firefox/</a>	Целевой страницей является страница авторитетного и известного сайта cneu.com, с которой можно загрузить браузер. Это полезная страница для транзакционного значения запроса.
		<a href="http://www.mozilla.com/en-US/firefox/firefox.html">http://www.mozilla.com/en-US/firefox/firefox.html</a>	Целевой страницей является страница с загрузкой браузера на официальном сайте Firefox. Эта страница может быть полезной для запросов со значением Do и Go.
[Nikon digital cameras], English (US)	Do, Know и Go. Данный запрос может быть транзакционным, информационным и навигационным. Пользователей может интересовать как покупка техники Nikon (транзакционный запрос), так и информация о бренде Nikon (информационный запрос) или ассортимент на странице сайта Nikon (навигационный запрос).	<a href="http://www.target.com/s/nikon+digital+cameras">http://www.target.com/s/nikon+digital+cameras</a>	Целевая страница - страница Nikon на сайте target.com. На данной странице представлено более 60 цифровых камер Nikon с ценами, описанием и обзорами.
		<a href="http://reviews.cnet.com/digital-camerareviews/?filter=1000036_108496_&amp;tag=centerColumnArea1.0">http://reviews.cnet.com/digital-camerareviews/?filter=1000036_108496_&amp;tag=centerColumnArea1.0</a>	Целевой страницей является страница с обзором цифровых камер Nikon (цены, типы камер, модели и свойства) на cnet.com. Данная страница позволяет пользователям сравнить цены, свойства техники и т.д. Данная страница является полезной для информационного значения запроса.
[ipad], English (US)	Do, Know и Go. Данный запрос может быть транзакционным, информационным и навигационным. Пользователей может интересовать как покупка iPad (транзакционный), так и информация о нем (информационный) или сама страница гаджета на сайте Apple.	<a href="http://www.engadget.com/2011/03/09/ipad-2-review/">http://www.engadget.com/2011/03/09/ipad-2-review/</a>	Целевой страницей является страница сайта engadget.com с подробным обзором iPad. Эта страница полезна для транзакционного значения запроса.
		<a href="http://www.apple.com/ipad/">http://www.apple.com/ipad/</a>	Целевой страницей является страница об iPad на официальном сайте Apple. Эта страница может быть целевой для данного запроса и полезной для информационного и навигационного значения данного запроса.
		<a href="http://store.apple.com/us/browse/home/shop_ipad/family/ipad?mco=OTY2ODA0NQ">http://store.apple.com/us/browse/home/shop_ipad/family/ipad?mco=OTY2ODA0NQ</a>	Целевой страницей является страница iPad в Apple Store. Здесь пользователь может и купить гаджет и найти необходимую информацию о нем. Эта страница может быть целевой для данного запроса и полезной для транзакционного, информационного и навигационного значения данного запроса.



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 3. ЯЗЫК ЦЕЛЕВОЙ СТРАНИЦЫ

Вам необходимо понимать выбранный вами язык задачи (язык, на котором вы будете оценивать выдачу) и английский язык. Также вы должны немного понимать языки, которые используются в вашем географическом регионе.

Страницам также нужно будет присваивать одну из следующих меток:

- **Язык задачи;**
- **Допустимый язык;**
- **Английский язык;**
- **Иностранный язык;**
- **Ни один из вышеперечисленного.**

**Язык задачи:** Используйте данную метку в том случае, если контент страницы полностью (или его большая часть) представлен на языке задачи.

**Допустимый язык:** В данном случае это языки, которые широко используются в выбранном вами регионе (местоположение задачи). Используйте данную метку в том случае, если контент страницы полностью (или его большая часть) представлен на допустимом языке.

**Английский язык:** Используйте данную метку в том случае, если контент страницы полностью (или его большая часть) представлен на английском языке.

**Иностранный язык:** Используйте эту метку в том случае, если вы считаете, что пользователи вашего региона НЕ смогут понять/прочитать контент страницы.

**Ни один из вышеперечисленного:** Используйте данную метку в том случае, если язык страницы невозможно определить. Например, на странице вообще нет текста или присутствуют только изображения. Или текст страницы сильно искажен, присутствуют ошибки в кодировке, не позволяющие идентифицировать язык.

При оценке страниц с несколькими языками руководствуйтесь здравым смыслом!

### 4. ОЦЕНКА ПО ШКАЛЕ РЕЛЕВАНТНОСТИ

Шкала релевантности страниц состоит из пяти оценок, основанных на намерении посетителя и полезности целевой страницы:

- **Витальная**
- **Полезная**
- **Релевантная**
- **Относительно релевантная**
- **Нерелевантная/неполезная**

В некоторых случаях используется такая категория, как **Нерейтингуемая**.

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

## 4.1 ВИТАЛЬНАЯ

Оценка **Витальная** используется в особых случаях:

- 1) По доминирующей интерпретации запрос предполагает навигацию/переход, цель такого запроса — конкретная страница.
- 2) По доминирующей интерпретации запрос предполагает определенный объект (такой как персона, место, компания, ресторан, продукт, организация и т.д.), целевой страницей должна быть официальная страница, связанная с данным объектом.

В обоих случаях это должна быть доминирующая интерпретация запроса! Если интерпретация запроса не является доминирующей, оценка **Витальная** не используется.

Большинство витальных страниц очень полезные. Однако помните, что это не является требованием к определению витальности. Некоторые витальные страницы официальны и не очень полезны.

### 4.1.1 ОСНОВНЫЕ ПОНЯТИЯ ВИТАЛЬНОСТИ

Для большинства запросов витальных страниц не существует. Вот некоторые примеры того, когда витальных страниц нет:

- у запроса нет доминирующей интерпретации,
- запрос не предполагает объект или навигацию,
- под запрос нет официального сайта или страницы,
- нет объекта, который может быть целью запроса.

Вот примеры ситуаций, когда под запрос нет витальной страницы:

Запрос	Витальная страница	Описание
[ADA], English (US)	не существует	Нет доминирующей интерпретации. Americans with Disabilities Act American Dental Association American Diabetes Association Все объекты данного запроса имеют общую интерпретацию. Под каждое значение запроса есть официальная страница, но ни одно из них не является доминирующим.
[knitting], English (US)	не существует	Это информационный запрос. Вязанием может заниматься любой, поэтому и сайтов на эту тему может быть очень много. Но официального ресурса о вязании быть не может.
[diabetes], English (US)	не существует	Это информационный запрос. Нет человека или организации, которые могли бы заявить о своем праве собственности на данную тему.
[ipod reviews], English (US)	не существует	[ipod] - объектный запрос, но [ipod reviews] - информационный запрос. Пользователи ищут информацию, которую могут предоставить многие сайты.
[how old is britney spears?], English (US)	не существует	[britney spears] - это объектный запрос, но [how old is britney spears?] таким не является. Это информационный запрос. Пользователи ищут информацию, которую могут предоставить многие сайты.



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 4.2 ПОЛЕЗНАЯ

Оценка **Полезная** ставится тем страницам, которые очень ценны для большинства пользователей. Такие страницы должны быть высокого качества и соответствовать запросу. Зачастую они обладают (полностью или частично) следующими характеристиками: удовлетворяющая запросу, авторитетная, занимательная и/или актуальная.

Полезные страницы хорошо структурированы, они внушают доверие. Обычно это страницы надежного сайта. На полезных страницах нет спама.

Обратите внимание, для одного запроса может быть несколько полезных страниц.

### 4.3 РЕЛЕВАНТНАЯ

Оценка **Релевантная** ставится тем страницам, которые могут быть полезными для многих или некоторых юзеров. Релевантные страницы имеют меньше определяющих характеристик, чем полезные. Так же как и полезные, релевантные страницы должны соответствовать запросу, однако они могут быть менее полными/исчерпывающими, менее актуальными, принадлежать менее авторитетным сайтам и освещать лишь один важный аспект запроса.

Релевантные страницы должны быть тематичны запросу и цепны для пользователя. Они не должны быть некачественными.

### 4.4 ОТНОСИТЕЛЬНО РЕЛЕВАНТНАЯ

Оценка **Относительно релевантная** ставится тем страницам, которые не очень полезны для большинства людей, но имеют некоторое отношение к запросу. **Относительно релевантные** страницы могут быть низкого качества и содержать меньше ценной для пользователя информации. **Относительно релевантные** страницы могут соответствовать второстепенной интерпретации запроса, иметь устаревшую информацию, быть очень специфичными или, наоборот, очень общими.

Оценка **Относительно релевантная** может ставиться мобильным версиям целевых страниц (относящихся к запросу), которые появляются в обычной выдаче. Страницы, предназначенные для мобильных пользователей, отличаются от обычных контентом (обычно его меньше) и функциональностью. Если мобильная целевая страница нерелевантна запросу, то ставится оценка **Нерелевантная/Неполезная**.

### 4.5 НЕРЕЛЕВАНТНАЯ/НЕПОЛЕЗНАЯ

Оценка **Нерелевантная/Неполезная** ставится тем страницам, которые могут быть полезны слишком малому количеству людей либо никому. **Нерелевантные/Неполезные** страницы не отвечают тематике запроса и/или не являются полезными.



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Также оценка **Нерелевантная/Неполезная** ставится в том случае, если страница не отвечает важному аспекту запроса, который определяет намерение пользователя.

Вам будут встречаться страницы, которые можно назвать абсолютно бесполезными (вводящими в заблуждение), они обязательно оцениваются как **Нерелевантные/Неполезные**. Например, страница с различными ссылками и рекламой и без какого-либо контента. Ссылки перенаправляют на другие страницы с подобными ссылками и рекламой. Когда вы не находите на странице ничего, что могло бы быть полезно юзеру, присваивайте странице оценку **Нерелевантная/Неполезная**. Такие страницы также выделяются меткой **Спам**, подробнее о которой вы узнаете из раздела 6.1 данной главы.

### 4.6 НЕРЕЙТИНГУЕМАЯ

Оценка **Нерейтингуемая** ставится тем страницам, которые невозможно оценить. Вы будете сталкиваться с разными типами нерейтингуемых страниц, поэтому используйте следующие категории для уточнения оценки:

- **Не загрузилась,**
- **Иностранный язык.**

Обратите внимание, страница может иметь несколько категорий. Например, если на целевой странице отображается сообщение об ошибке на иностранном языке и больше нет никакого контента, то данная страница оценивается следующим образом: **Нерейтингуемая: Не загрузилась** и **Нерейтингуемая: Иностранный язык**.

#### 4.6.1 НЕРЕЙТИНГУЕМАЯ: НЕ ЗАГРУЗИЛАСЬ

**Нерейтингуемая: Не загрузилась** (или просто **Не загрузилась**) — особая оценочная категория для страниц, которые не загружаются или не имеют никакого контента. Обычно такие страницы содержат или сообщение о какой-либо ошибке, или ничего.

Страницы, относящиеся к категории **Нерейтингуемая: Не загрузилась**, — это:

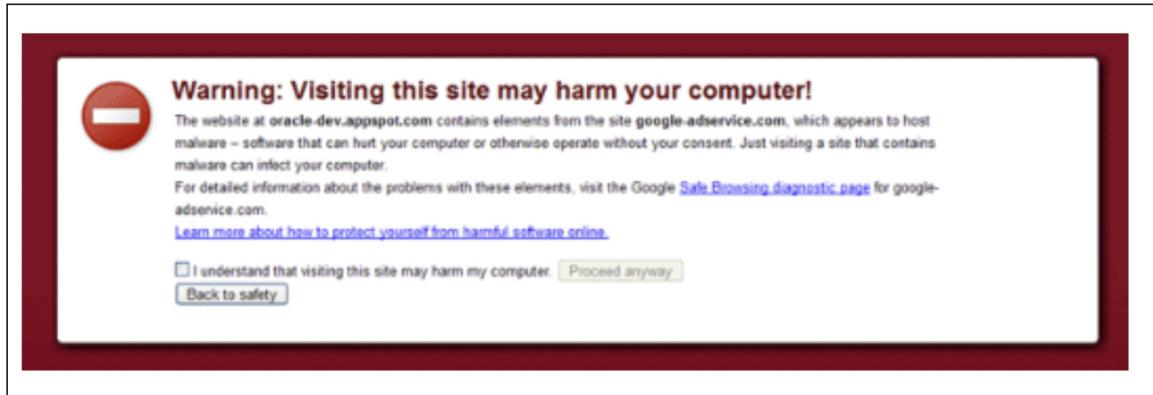
- Страницы с сообщениями об ошибке и без какого-либо другого контента.
- Страницы с неработающими редиректами и без какого-либо другого контента.
- Абсолютно пустые страницы.
- Страницы с предупреждениями: «Внимание — посещение данного сайта может навредить вашему компьютеру».
- Страницы с запросом на сертификат.

Помните, вы не должны отмечать страницу меткой **Спам** или **Вредоносная** только потому, что на ней появилось окошко с предупреждением или запрос на сертификат. На хороших страницах такое тоже встречается. Поэтому не ставьте метки **Спам** или **Вредоносная**, если для этого нет других причин.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Подробное описание меток **Спам** и **Вредоносная** страница вы найдете в разделе 6.1 и 6.3 данной главы.

Пример предупреждающего сообщения:



Пример запроса на сертификат:



Здесь [http://en.wikipedia.org/wiki/List\\_of\\_HTTP\\_status\\_codes](http://en.wikipedia.org/wiki/List_of_HTTP_status_codes) вы найдете описание различных видов сообщений об ошибках. Наиболее распространенными являются:

- 401 Unauthorized — для доступа к запрашиваемому ресурсу требуется аутентификация.
- 403 Forbidden — сервер отказывается выполнять запрос из-за ограничений в доступе для клиента к указанному ресурсу.
- 404 Not Found — сервер не нашёл соответствующего ресурса.
- 500 Internal Server Error — любая внутренняя ошибка сервера.
- 503 Service Unavailable — сервер временно не имеет возможности обрабатывать запросы по техническим причинам.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Страницы, которые загружаются не полностью или имеют нерабочие ссылки, должны оцениваться в соответствии с их полезностью.

Примеры страниц с ошибками (и без другого контента), которые должны быть отмечены как **Не загрузилась**:

Запрос	Ошибка на целевой странице	Оценка	Пояснение
[Douglas Instruments], English (US)	404 Not Found. Запрашиваемая вами страница не найдена на сервере	Не загрузилась	Страница, отдающая такое сообщение и не имеющая контента, должна оцениваться как <b>Не загрузилась</b>
[SIAD], English (US)	У вас недостаточно прав для просмотра данной страницы.	Не загрузилась	Страница, отдающая 403 ошибку и не имеющая контента, должна оцениваться как <b>Не загрузилась</b>
[electionwatch 2009.com], English (US)	Осторожно - посещение данного сайта может навредить вашему компьютеру!	Не загрузилась	Страница с таким сообщением должна оцениваться как <b>Не загрузилась</b>
[hat shipping], English (US)	Сайт находится в разработке	Не загрузилась	Пустая страница с фразой "Сайт находится в разработке" оценивается как <b>Не загрузилась</b>

Обратите внимание, что целевая страница с сообщением об ошибке, но с контентом и рабочими ссылками должна оцениваться сообразно их полезности. Сообщения об ошибке на таких страницах обычно настроены веб-мастерами. Но утверждать, что это действительно так, можно не всегда. Поэтому важно изучить контент и/или ссылки на странице.

Помните, иногда на страницах с сообщением об ошибке присутствуют ссылки и текст, которые могут быть ошибочно приняты за контент. Скорее всего, это автоматически генерированное сообщение.

Когда вы ставите оценку **Нерейтингуемая: Не загрузилась**, пожалуйста, скопируйте и вставьте сообщение об ошибке, которое встретилось вам на целевой странице, в блок с комментариями.

### ВЫБОР ЯЗЫКА ЦЕЛЕВОЙ СТРАНИЦЫ ДЛЯ СТРАНИЦ, КОТОРЫЕ НЕ ЗАГРУЖАЮТСЯ

**Вам будет нужно выбрать метки для определения языка целевой страницы под каждое оценочное задание, даже если страница не загружается:**

- Используйте метку **Язык задачи** для страниц на том языке, на котором вы работаете.
- Используйте метку **Допустимый язык** для страниц на языках вашего региона.
- Используйте метку **Английский язык** для страниц на английском.
- Используйте метку **Иностранный язык** для страниц на языке, отличном от языка задачи, допустимого языка или английского.
- Используйте метку **Ни один из вышеперечисленных** для пустых страниц, для страниц без привязки к какому-либо языку и для страниц, которые не загружаются.

Описание меток для определения языка целевой страницы вы найдете в разделе 3 данной главы.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 4.6.2 НЕРЕЙТИНГУЕМАЯ: ИНОСТРАННЫЙ ЯЗЫК

Оценка **Нерейтингуемая: Иностранный язык** ставится в том случае, если язык целевой страницы отличается от языка задачи, допустимого языка или английского.

Оценку **Нерейтингуемая: Иностранный язык** вы будете применять всегда, когда используете метку **Иностранный язык** при определении языка целевой страницы.

Оценка **Нерейтингуемая: Иностранный язык** подходит для всех типов запросов и всех страниц на иностранных языках, даже если вы понимаете язык страницы и считаете данную страницу нерелевантной.

Также ставьте оценку **Нерейтингуемая: Иностранный язык** в тех случаях, когда сомневаетесь.

## 5.0 ОТ НАМЕРЕНИЯ ПОЛЬЗОВАТЕЛЯ ДО ВЫСТАВЛЕНИЯ ОЦЕНКИ

В предыдущих разделах данного руководства вы уже встречались с понятиями «запросы» и «шкала релевантности». В данном разделе мы объединим их. Наиболее важными факторами, которые нужно принимать во внимание при оценке, являются намерение пользователя и полезность страницы.

Другие важные идеи раздела:

- Вы должны представлять пользователя вашего региона. Вы должны оценивать страницы с точки зрения пользователя.
- У некоторых запросов множество интерпретаций и пользовательских намерений. Это не повод занижать оценку.
- Асессоры отличаются от пользователей. Результаты, полезные для асессоров, не обязательно имеют ценность для пользователей.
- Местоположение играет важную роль. Хорошие страницы должны соответствовать запросу географически.

## 5.1 НАМЕРЕНИЕ ПОЛЬЗОВАТЕЛЯ И ПОЛЕЗНОСТЬ СТРАНИЦЫ

Очень важно понимать намерение пользователя, т.к. вы будете оценивать целевые страницы по тому, насколько хорошо они соответствуют намерению пользователя. Для этого вам понадобятся:

- ваш поисковый опыт,
- ваш здравый смысл,
- онлайн-исследования.

Если вы считаете, что страница не имеет ценности для пользователя, ставьте низкую оценку. Релевантная страница должна быть полезна для человека. Относительно релевантная страница менее полезна, но тематична запросу. Нерелевантная не несет никакой пользы и не соответствует тематике запроса.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Не «вымучивайте» оценки! Оценивайте страницу и двигайтесь дальше. Если вы выбираете между двумя оценками — выбирайте меньшую. Если из трех — действуйте на свое усмотрение.

Оценивая страницу, иногда полезно обращать внимание на ее URL.

Вот в каких именно ситуациях:

- Для определения спама.
- Для обнаружения редиректов.
- Для определения витальных страниц.

**Помните, вы должны ВСЕГДА переходить на целевую страницу.**

### 5.2 О ВАЖНОСТИ МЕСТОПОЛОЖЕНИЯ

Хорошие поисковые системы в нужных случаях предлагают пользователю местные/региональные результаты, т.е. страницы, отвечающие местоположению пользователя. Например, если пользователь из США ищет пиццу, то вряд ли его интересуют пиццерии в Лондоне.

**Важно:** несмотря на содержание запроса, мы считаем, что большинству пользователей важно получить локальные ответы.

В большинстве случаев вам будет нужно занижать оценку, если контент страницы не относится к нужному региону. Не стесняйтесь ставить странице оценку **Нерелевантная**, если ее содержание не отвечает географическим требованиям запроса, что делает ее бесполезной для посетителя.

### 5.3 О ВАЖНОСТИ ЯЗЫКА (для ассессоров, работающих не с английским языком)

Если ваш язык задачи — английский (English (US), English (UK), English (CA)), то вы можете пропустить этот раздел.

Достаточно часто вы будете использовать оценку **Нерейтингуемая: Иностранный язык** для тех случаев, когда язык страницы не соответствует языку задачи/английскому/допустимому языку. Подробно об этом рассказано в разделе 4.1.5 данной главы.

Целевые страницы на языке запроса — всегда лучший выбор для пользователя.

Даже если целевые страницы не на иностранном языке, а на английском или допустимом языке, они не всегда могут соответствовать запросу пользователя географически. Например, в некоторых странах уровень знания английского языка достаточно высок. В таких государствах страницы на английском в выдаче будут вполне релевантны. Но в странах, где уровень знания английского ниже, страницы на данном языке могут оказаться неполезными для посетителей.

Отметим, что некоторые запросы «требуют» результатов на английском или допустимом языке. Ваша оценка страницы на английском или допустимом языке должна основываться на ее полезности для юзера. Помните, что страницам, которые не очень полезны для большинства людей, но всё же имеют отношение к запросу, нужно ставить оценку **Относительно релевантные**.



Понравилась книга? Расскажи друзьям!

# ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

В некоторых странах английский язык является одним из официальных языков или широко распространенным. Пользователи из этих стран не расстроются, если целевые страницы под их запросы будут на английском языке. Например, в Сингапуре четыре официальных языка: английский, малайский, китайский и тамильский. Наибольшее распространение имеет английский.

## 5.4 МНОГОЗНАЧНЫЕ ЗАПРОСЫ

Вам придется оценивать целевые страницы под запросы, обладающие несколькими значениями и подразумевающие несколько намерений пользователя.

- Страницы под запросы с второстепенной интерпретацией должны оцениваться ниже других.
- Страницы под запросы с общей интерпретацией не должны оцениваться низко.
- Только страницы под запросы с доминирующей интерпретацией могут оцениваться как **Витальные**.

Вот несколько примеров:

Интерпретация запроса	Пример	Оценка
<b>Доминирующее значение:</b> из всех пользователей, которые вводят данный запрос, большинство предполагает данное значение	[apple], English (US): Компьютеры Apple. Большинство пользователей, вводящих данный запрос, ищут именно компьютеры Apple.  [windows], English, (US): Microsoft ОС. Большинство пользователей, вводящих данный запрос, ищут именно эту операционную систему.  [amazon], English (US): Популярный сайт amazon.com. Большинство пользователей, вводящих данный запрос, хотят перейти на этот сайт.  [median], English (US): Математический термин. Большинство пользователей, вводящих данный запрос, ждут информации о медиане. Несмотря на то, что у этого запроса есть доминирующая интерпретация, ставить оценку <b>Витальная</b> нельзя, т.к. данная информация никому не принадлежит. Максимально возможная оценка - <b>Полезная</b> .  [guinea pig], English (US): Морская свинка. Большинство пользователей, вводящих этот запрос, ищут информацию о данном зверьке. Несмотря на то, что у запроса есть доминирующая интерпретация, ставить оценку <b>Витальная</b> нельзя, т.к. данная информация никому не принадлежит. Максимально возможная оценка - <b>Полезная</b> .	от <b>Витальной</b> до <b>Нерелевантной/Неполезной</b>

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Интерпретация запроса	Пример	Оценка
<b>Общее значение:</b> из всех пользователей, которые вводят данный запрос, многие или некоторые предполагают именно это значение	<p>[apple], English (US): Фрукт. Некоторые пользователи, вводящие данный запрос, ищут информацию о фрукте.</p> <p>[windows], English (US): Окна. Многие или некоторые пользователи, вводящие данный запрос, ищут информацию про окна.</p> <p>[amazon], English (US): Тропические леса или река в Южной Америке. Некоторые пользователи, вводящие данный запрос, ищут информацию о тропических лесах или реке.</p> <p>[ada], English (US): American Dental Association (Американская стоматологическая ассоциация), American Diabetes Association (Американская ассоциация диабета) или American with Disabilities Act (Закон об американцах-инвалидах). Многие или некоторые пользователи, вводящие данный запрос, могут искать информацию о любой из этих тем.</p> <p>[mercury], English, (US): Автомобильный бренд, планета или химический элемент. Многие или некоторые пользователи, вводящие данный запрос, могут искать любое из этих значений.</p>	<b>от Полезной до Нерелевантной/Неполезной</b> <i>Для запроса с общей интерпретацией не может быть Витальных страниц</i>
<b>Второстепенное значение:</b> из всех пользователей, которые вводят данный запрос, лишь немногие предполагают это значение	<p>[ada], English (US): Atlanta Development Authority или American Darters Association. Лишь немногие пользователи ищут данные значения запроса.</p> <p>[mercury], English (US): Mercury Magazine (журнал) или Mercury Records (компания звукозаписи). Лишь немногие пользователи ищут данные значения запроса.</p> <p>[hot dog], English (US): Hot Dog - фильм 1984 года. Лишь немногие пользователи ищут данное значение запроса.</p>	<b>от Релевантной до Нерелевантной/Неполезной</b>
<b>"Без шансов":</b> практически никто не предполагает данное значение	[guinea pig], English (US): Свинья, живущая в Новой Гвинеи. Маловероятно, что кто-то будет искать информацию о свиньях с Новой Гвинеи. Скорее, пользователь ищет морскую свинку.	<b>Нерелевантная/Неполезная</b>

Помните, что запросы с доминирующей интерпретацией могут также иметь и общие интерпретации.

Запрос	Доминирующая интерпретация	Общая интерпретация
[windows], English (US)	Операционная система Microsoft	Окна
[kayak], English (US)	Туристический сайт	Байдарка/каяк



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Помимо многозначности запросов, вы будете сталкиваться с многозначностью пользовательских намерений. Вам нужно будет определить, насколько намерение пользователя является вероятным/определенным. Чем намерение менее вероятно, тем меньше оценка.

Намерение пользователя	Примеры	Оценка
<b>Вероятное намерение:</b> У многих или большинства такие намерения.	[tetris], English (US): играть в видеоигру онлайн или загрузить. [flowers], English (US): заказать цветы онлайн, найти информацию или изображения цветов. [credit cards], English (US): найти банки, запросить оформление кредитной карты, сравнить кредитки разных банков. [amazon], English (US): перейти на сайт amazon.com.	от <b>Витальной</b> до <b>Нерелевантной/Неполезной</b>
<b>Менее вероятное намерение:</b> У немногих или некоторых такие намерения.	[tetris], English (US): изучить историю игры "Тетрис". [flowers], English (US): найти определение слова "цветы". [credit cards], English (US): прочитать статью об истории кредитных карт. [amazon], English (US): прочитать обзорную статью об Amazon.	от <b>Релевантной</b> до <b>Нерелевантной/Неполезной</b> Оценка базируется на том, скольким пользователям данная страница будет полезна.

### 5.5 ЧЕТКОСТЬ ЗАПРОСОВ И ЦЕЛЕВЫХ СТРАНИЦ

Запросы могут быть четкими и нечеткими или же иметь ту или иную степень четкости. Вот некоторые примеры:

Запрос	Более четкий запрос	Еще более четкий запрос
[стул]	[стул для столовой]	[ikea "henriksdal" мягкий стул с высокой спинкой]
[фотоаппарат]	[фотоаппарат Nikon]	[Nikon d5000 slr]
[Toyota]	[Toyota гибрид]	[Toyota Prius 2010]
[библиотека]	[Гарвардская библиотека]	[Гарвардская астрологическая библиотека]

Хорошая целевая страница должна соответствовать степени четкости запроса, чтобы быть полезной. Если целевая страница не полностью отвечает запросу, вы должны внимательно подойти к ее оценке — определить, насколько она может быть ценной для пользователя.



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

### 5.6 РАСПРОСТРАНЕННЫЕ ОШИБКИ ПРИ ОЦЕНКЕ СТРАНИЦ

Ниже приведены распространенные ошибки, которые допускают асессоры при оценке целевых страниц. Большинство ошибок связано с определением намерения пользователя и соответствия целевой страницы запросу.

#### 5.6.1 СТРАНИЦЫ ИЗ СЛОВАРЕЙ И ЭНЦИКЛОПЕДИЙ

Страницы из словарей и энциклопедий очень часто полезны асессорам, пытающимся вникнуть в суть запроса. Иногда они могут быть полезны и юзерам, но не в том случае, если пользователь уже знает смысл слов в запросе и ищет что-то другое.

Асессоры часто ставят высокую оценку страницам, отвечающим информационным намерениям пользователя, даже если запрос транзакционный. Однако если запрос явно транзакционный, то максимальной оценкой для информационных страниц может быть **Релевантная**. Всегда задавайте себе вопрос: пользователь хочет что-то узнать или что-то сделать? В соответствии с вашим ответом оценивайте целевую страницу.

Обратите внимание на информационное намерение такого класса запросов, как «Товары». Многие пользователи перед покупкой изучают товар в сети. Поэтому запросы класса «Товары» относятся как к информационным запросам, так и к транзакционным.

#### 5.6.2 ЗАПРОСЫ С ОШИБКАМИ И ОПЕЧАТКАМИ

Вы будете встречать запросы, написанные с ошибками или опечатками. В случаях, когда запрос явно написан с ошибкой, ваша оценка опять же должна определяться намерением пользователя без внимания на то, как он ввел запрос.

Для запросов с неочевидной ошибкой/опечаткой вы должны представить себя на месте пользователя, который вводит такой запрос.

Например, по запросу [federal expres], English (US) пользователь явно ищет Federal Express. В запросе [my sapce], English (US), скорее всего, опечатка, и пользователю нужна соцсеть MySpace. Для этих запросов правильный вариант вполне очевиден.

А теперь представьте такой запрос [John Stuart], English (US). Ассесор может решить, что пользователь ищет страницы о Джоне Стюарте, известном комедийном актере и ведущем шоу The Daily Show. Запрос написан правильно. Но есть еще (как минимум) один человек по имени Джон Стюарт (он продюсер шоу в Лас-Вегасе), и его имя полностью совпадает с запросом пользователя. Принимайте это во внимание.

Не считайте запрос неправильно написанным в том случае, если он совпадает с именем человека или названием организации. Предположите, что такой человек мог бы существовать. Но бывает и так, что для запроса не существует результатов.



Понравилась книга? Расскажи друзьям!

## 5.6.3 URL-ЗАПРОСЫ

Некоторые запросы имеют вид ссылки, это URL-запросы. Такие запросы могут быть по-разному введены пользователем, например, с www и без или с http://. Запросы типа [www.ibm.com], [http://www.ibm.com], [ibm.com] считаются одинаковыми.

Некоторые запросы состоят из названия сайта/сервиса, например, [yahoo] или [yahoo mail]. Они не содержат таких компонентов ссылки, как «www» или «.com», но все равно считаются URL-запросами.

Но большинство запросов не является ни URL-запросами, ни запросами с названиями сайтов. Очень часто компоненты запроса не относятся к определенному сайту/странице.

Вот несколько примеров:

URL-запросы	Запросы с названием сайта/страницы <i>(это навигационные запросы)</i>	Общие запросы
[ebay.ca] [amazon.com] [people.com] [bbc.co.uk] [www.dealbook.com] [mail.yahoo.com] [news google.com] [tax form 1040 irs.gov] [rei.com]	[ebay] [amazon] [people] [bbc] [dealbook] [yahoo mail] [google news] [irs 1040 tax form official page] [rei kayak page]	[couches] [diabetes] [weight loss] [tax forms] [quilting]

Начнем с URL-запросов. Некоторые из них представлены нерабочими ссылками. Т.е., если вы скопируете эту ссылку в адресную строку, она не откроется. Однако мы считаем, что пользователю нужна именно эта страница. Такие запросы мы называем дефектными URL-запросами. Существует много типов дефектных URL-запросов. Вот описание некоторых из них:

- Ссылка в запросе имеет правильный вид, но страница не загружается. Например, [www.UnitedStatesPassportProvider.com].
  - Ссылка в запросе имеет правильный вид, но, очевидно, закралась опечатка, и ссылка не открывается. Например, [www.mcriosoft.com].
  - Запрос похож на URL-формат, но имеет лишние слова и/или пробелы. Например, [Australian open tennis tournament.com].
  - Запрос состоит из слов и ссылок. Например, [barbie.com dress up games].

Некоторые URL-запросы достаточно сложно оценивать. В любом случае вам нужно перейти на целевую страницу и оценить ее контент. Также обращайте внимание на адрес целевой страницы и URL в запросе, сравнивайте их. Не оценивайте URL-запросы только по внешнему виду ссылки.

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Интерпретировать намерение пользователя для дефектных URL-запросов достаточно сложно, т.к. пользователь легко мог опечататься.

Если URL-запрос имеет правильный вид и рабочую ссылку, значит, доминирующей интерпретацией запроса будет страница с данным URL. Странице можно присвоить оценку **Витальная** в том случае, если ее адрес полностью совпадает с URL, указанным в запросе. Обратите внимание, иногда URL целевой страницы может быть несколько длиннее указанного в запросе или немного отличаться. Например, для запроса [imagehosting.us] обе ссылки <http://imagehosting.us/> и <http://www.imagehosting.us/index.php?action=show&ident=1478428> должны быть оценены как **Витальные**, т.к. целевая страница одна.

Если URL-запрос имеет неправильный вид и нерабочую ссылку, определяйте намерение пользователя на ваше усмотрение. Но не ставьте оценку **Витальная**, даже если у вас нет сомнений, что страница целиком отвечает намерению пользователя.

Перейдем к запросам с названием сайта/страницы. Это запросы, которые состоят из названия сайта или страницы, доминирующей интерпретацией такого запроса является сайт/страница. Однако некоторые запросы с названием сайта/страницы могут предполагать не только сайт/страницу.

Запрос с названием сайта/страницы	Объяснение
[kayak], English (US)	Пользователь мог бы искать байдарку, но Kayak - очень популярный туристический сайт. Поэтому доминирующая интерпретация - сайт kayak.com.
[youtube], English (US)	YouTube - один из самых популярных ресурсов в сети
[ebay], English (US)	eBay - один из самых популярных ресурсов в сети
[webmd], English (US)	Webmd - один из самых популярных ресурсов на медицинскую тему в сети
[twitter], English (US)	Twitter - популярный веб-ресурс
[cafepress], English (US)	Cafepress - сайт, на котором пользователи могут купить футболки и другие подарки
[addicting games], English (US)	AddictingGame - очень популярный игровой сайт
[rei kayak page], English (US)	Пользователь хочет перейти на страницу о байдарках на сайте Rei



Понравилась книга? Расскажи друзьям!

## ГЛАВА I: ПРИНЦИПЫ ОЦЕНКИ

Ниже несколько примеров запросов, которые не содержат название сайта и URL. В запросе присутствует название, и оно может совпадать с названием сайта, но пользователь ищет неодноименный ресурс. У таких запросов нет витальных страниц.

Общий запрос	Объяснение
[birdcages], English (US)	Вероятно, пользователь изучает или планирует купить клетку для птиц. Это общий запрос. Витальной страницы нет. Существует сайт с таким же названием birdcages.com, но клетки для птиц продают многие другие.
[kamasutra], English (US)	Вероятно, пользователь хочет узнать больше о «Камасутре» или прочитать текст. Витальной страницы нет. Существует сайт с названием kamasutra.com, но, скорее всего, это не доминирующая интерпретация запроса.
[weightloss], English (US)	Пользователь ищет информацию о снижении веса. На эту тему в сети много авторитетных страниц. Есть одноименный сайт weightloss.com с полезной информацией по снижению веса. Но, вероятней всего, у пользователя нет намерения перейти туда.
[couches], English (US)	Пользователь ищет информацию или планирует купить диван. В сети много хороших сайтов, где можно купить диваны. Существует сайт couches.com, но в запросе ничего не говорит о том, что пользователь ищет именно этот сайт.

Помните, что любой вопрос может легко превратиться в URL-запрос простым добавлением «.com» к слову. Но, если в запросе нет «.com», вы не должны думать, что слово в запросе является названием сайта.

Другими словами, т.к. запрос звучит просто [диваны], это означает, что сайт <http://www.couches.com> — не то, что ищет пользователь. Будьте очень внимательны с общими запросами. Широко распространена такая спам-технология, как создание сайтов с общими доменными именами.

Когда пользователи вводят URL-запросы, они хотят перейти на соответствующую страницу. Такая страница должна быть оценена как **Витальная**. Отметим, что намного сложнее для таких запросов определять и оценивать страницы как «не Витальные». Иногда витальная страница — единственный полезный результат для URL-запроса. Но порой и другие страницы могут быть полезными, возможная оценка таких страниц варьируется от **Полезной** до **Нерелевантной/Неполезной**.

## 5.6.4 НОВЫЕ И СТАРЫЕ СТРАНИЦЫ

Информационные, или know-, запросы могут быть посвящены как актуальному событию, так и уже прошедшему ранее. Целевая страница должна быть оценена на основе того, насколько она отвечает информационным требованиям запроса. На некоторые запросы требуются очень актуальные результаты. Правда, чаще вам будет нужно анализировать контент страницы, а не дату его публикации.

Для некоторых запросов очень важна актуальность и свежесть результатов. Для запросов о событиях, недавно случившихся или еще продолжающихся, нужны страницы с актуальными сведениями. Если пользователь ищет результаты выборов, спортивные мероприятия или другие виды ежегодных соревнований, то ему важны самые свежие результаты, а не прошлогодние.

В зависимости от того, когда происходит то или иное ежегодное мероприятие, наиболее полезная страница может быть посвящена его прошедшим/текущим или будущим событиям. Например, если мероприятие прошло несколько месяцев назад, то наиболее полезная страница будет с контентом о прошедшем событии. И наоборот, если мероприятие только планируется, полезная страница должна рассказывать о нем.

Если целевая страница оказалась официальной страницей мероприятия, ее нужно оценивать как **Витальную**, даже если контент на ней посвящен прошедшему или грядущему мероприятию.

А вот информационные запросы могут требовать очень свежих результатов. Например, запрос [население Парижа], пользователь ищет самые последние статистические данные.

С другой стороны, если запрос звучит как [население Франции в 1813 году], то свежесть данных на странице не имеет столь большого значения, как сама запрашиваемая информация. Иногда «старые» страницы являются единственным качественным источником информации о прошедших событиях. Помните, «старые» страницы не значит «устаревшие» или плохие. Всё зависит от запроса и контента.

Если вам для оценки досталась пустая страница результатов поиска с пустой поисковой строкой, она оценивается как **Нерелевантная** или **Неполезная**.

Если на странице присутствуют результаты поиска, она полезна. В зависимости от того, насколько полезна страница, ей могут быть поставлены оценки от **Полезная** до **Нерелевантная/Неполезная**.

## 6. МЕТКИ

Помимо выставления оценок, вам нужно будет присваивать специальные метки разным типам страниц.

## 6.1 МЕТКА СПАМ

Поводом для присвоения странице метки **Спам** являются специальные спам-сигналы, о которых будет подробно рассказано в главе 4 данного руководства.

**Не Спам (Not Spam):** если вы считаете, что страница создана без использования спам-методов, то присваивайте ей метку **Не Спам**.

**Возможно, спам (Maybe Spam):** если страница кажется вам спамной, но вы сомневаетесь в этом, следует присвоить ей метку **Возможно, спам**.

**Спам (Spam):** если вы считаете, что страница создана с использованием спам-методов, описанных в данном руководстве, присваивайте ей метку **Спам**.

Если вы присваиваете странице сразу две метки **Возможно, спам** и **Спам**, вам следует пояснить почему.

## 6.2 МЕТКА ПОРНО

Присваивайте данную метку всем страницам, содержащим контент порнографического характера (порнокартинки, ссылки, текст, pop-up и другая реклама и т.д.). Но обратите внимание на культурные особенности различных стран: в одних странах картинка может считаться изображением порнографического содержания, в других — нет.

### 6.2.1 НЕОЧЕВИДНЫЕ ПОРНОЗАПРОСЫ

Если пользователь вводит очевидный не порнозапрос, а в ответ получает целевую страницу с порноконтентом, такая страница должна оцениваться как **Нерелевантная/Неполезная** и получать метку **Порно**. Например, запрос [фото автомобилей]. Если вам выдается страница с обнаженной девушкой на машине, ставьте оценку **Нерелевантная/Неполезная** и метку **Порно**. Даже несмотря на то, что автомобиль на картинке присутствует.

### 6.2.2 ВОЗМОЖНЫЕ ПОРНОЗАПРОСЫ

Некоторые запросы имеют порно- и не порноинтерпретации. Например, в следующих англоязычных запросах могут иметь место порнонамерения: [girls], [gay], [thong], [breast], [sex], [spanking]. Такие запросы называются возможными порнозапросами (Possible Porn Inten).

Помните, для подобных запросов доминирующим намерением должно считаться не порнонамерение. Например, запрос [отшлётать] в первую очередь означает «воспитательное действие родителей по отношению к детям» (не порноинтерпретация).

Оценивайте порноинтерпретации запроса как второстепенные, даже если вы понимаете, что пользователь ищет порноконтент.

### 6.2.3 ОЧЕВИДНЫЕ ПОРНОЗАПРОСЫ

Целевые страницы для очевидных порнозапросов оцениваются по шкале релевантности, как и другие страницы, без занижения оценки. В дополнение странице ставится метка **Порно**.

Помните, вы не должны просто оценивать все порностраницы под порнозапросы как **Релевантные** или **Полезные**. Даже если это выдача по порнозапросу, целевая страница должна подходить под запрос и отвечать ожиданиям пользователя.

Страницам с вредоносным ПО ставится низкая оценка, даже если на них есть релевантный контент.

Такие запросы, как порнозвёзды, порнофильмы, порносайты и т.д. могут иметь витальные страницы. Но всегда помните, что и витальные страницы (в данном случае) заслуживают метку **Порно**.

Не присваивайте метку **Порно** странице, на которой нет порноконтента, но она была выдана по порнозапросу. Если целевая страница не содержит порноконтент, метка **Порно** не ставится.

### 6.3 МЕТКА ВРЕДОНОСНАЯ СТРАНИЦА

Странице присваивается данная метка, если:

- Вы вынуждены закрыть браузер из-за всплывающих окон, которые не закрываются.
- Были попытки загрузки вредоносных программ.

Обратите внимание, что pop-up реклама, которую вы можете закрыть, не является вредоносной.

Не присваивайте данную метку тем страницам, на которых вам встретились предупреждающие сообщения или запросы на сертификат. Метка **Вредоносная** присваивается только в описанных выше случаях. Страницам, на которых появилось окошко с предупреждением о вредоносных программах, метка **Вредоносная** не присваивается. Такие страницы получают оценку **Нерейтингуемая: Не загрузилась**.

### 6.4 СОВМЕСТИМОСТЬ ОЦЕНОК И МЕТОК

Помните, страницам с оценкой **Нерейтингуемые** могут быть присвоены следующие метки: **Спам**, **Порно** и/или **Вредоносная страница**. Вот несколько примеров:

- Страница на иностранном языке с порнокартинками.
- Страница на иностранном языке со скрытым текстом.
- Страница не загружается, но по URL можно определить скрытый редирект. Подробно о скрытом редиректе будет рассказано в главе 4 данного руководства.
  - Страница не загружается и содержит порнорекламу.
  - Страница на иностранном языке, вы не можете закрыть pop-up рекламу на ней и вынуждены уйти со страницы.

## ГЛАВА II: ОЦЕНКА URL С УЧЕТОМ РЕГИОНА ЗАПРОСА

### 1. РЕГИОН ЗАПРОСА

Как мы писали ранее, оценивая выдачу, асессоры также принимают во внимание местоположение задачи — рабочий регион асессора, которым обычно является страна.

Помимо этого, вы будете сталкиваться с «регионом запроса», который отражает географическое местонахождение пользователя, вводящего запрос. Регион запроса могут определять почтовый индекс, название города или деревни, области и т.п. Обычно регион запроса определяется поисковыми системами автоматически, но может сообщаться в самом запросе пользователя.

Для узкоопределенного региона запроса (ограниченного почтовым индексом, названием города и т.д.) могут подойти результаты с более широкой локацией. Помните, что пользователи могут искать ближайший к ним магазин или ресторан. Но человека также могут устроить расширенные результаты — с объектами, находящимися за пределами заданной географической зоны. Подобные случаи вы будете оценивать на свое усмотрение.

Важные моменты для понимания особенностей региона запроса:

- Вам нужно оценивать с точки зрения человека, живущего в зоне запроса.
- Локальные страницы (страницы, связанные с регионом запроса), которые могут быть полезны посетителю, должны получить высокую оценку.
- Страницы, полезные посетителям вне зависимости от региона запроса, также должны получить высокие оценки.
- Если запросом является некий объект (бизнес, организация, школа и т.д.), у которого есть и официальная главная, и локальная версия страницы, то обеим страницам присваивается оценка **Витальная** (Appropriate Vital).

**Важно:** Иногда пользователи уточняют регион в самом запросе. Например, в запросе [Магнит, «Новомосковск, Тульская область»] пользователь указал Новомосковск, Тульская область как исключительное местоположение. Обратите внимание, что для некоторых запросов свойственны и регион запроса (запрос сделан из определенной локации), и геопривязка в самом запросе (локация, запрашиваемая пользователем). В таких случаях страница оценивается относительно локации, указанной в запросе.

Три типа задач:

- Запрос, содержащий геопривязку.
- Запрос с регионом запроса/пользователя.
- Запрос с регионом пользователя и геопривязкой.

## ГЛАВА II: ОЦЕНКА URL С УЧЕТОМ РЕГИОНА ЗАПРОСА

	Скриншоты		Описание
<b>Это геонезависимый запрос, т.к. не указан регион запроса.</b>  <b>Обратите внимание, что в запросе регион уточнен.</b>	Запрос	pizza hut <i>san francisco</i>	Пользователь ищет информацию о Pizza Hut в регионе Сан-Франциско.
	URL	<a href="http://www.yelp.com/biz/pizzahut-san-francisco">http://www.yelp.com/biz/pizzahut-san-francisco</a>	
	Местоположение задачи	США	
	Язык задачи	Английский	
	Др. приемлемые языки	Нет	
<b>Это геозависимый запрос, т.к. присутствует регион запроса</b>	Запрос	pizza hut	Запрос был введен пользователем из Сан-Франциско. Можно предположить, что пользователь ищет пиццерию в Сан-Франциско.
	Регион запроса	***Сан Франциско***	
	URL	<a href="http://www.yelp.com/biz/pizzahut-san-francisco">http://www.yelp.com/biz/pizzahut-san-francisco</a>	
	Местоположение задачи	США	
	Язык задачи	Английский	
	Др. приемлемые языки	Нет	
<b>Это тоже геозависимый запрос, т.к. указан регион запроса.</b>  <b>Обратите внимание, что в запросе регион уточнен.</b>	Запрос	pizza hut <i>san francisco</i>	Запрос введен пользователем из Нью-Йорка. Но, т.к. в запросе присутствует "сан-франциско", пользователь ищет пиццерии в данном регионе. Несмотря на то, что регион запроса - Нью-Йорк.
	Регион запроса	***Нью Йорк***	
	URL	<a href="http://www.yelp.com/biz/pizzahut-san-francisco">http://www.yelp.com/biz/pizzahut-san-francisco</a>	
	Местоположение задачи	США	
	Язык задачи	Английский	
	Др. приемлемые языки	Нет	

## ГЛАВА II: ОЦЕНКА URL С УЧЕТОМ РЕГИОНА ЗАПРОСА

### 2. ОЦЕНКА СТРАНИЦЫ ПОД ГЕОЗАВИСИМЫЙ ЗАПРОС

Оценка страницы под геозависимый запрос аналогична оценке страницы под обычный запрос. За исключением того, что она содержит дополнительную информацию согласно региону запроса.

Информация	Оценка страницы под обычный запрос	Оценка страницы под геозависимый запрос
Регион запроса	На странице нет данной информации	***Нью-Йорк*** ***90210*** ***Даллас, Штат Техас*** ***Штат Техас***

#### Оценка страницы под геозависимый запрос

Дом. страница асессора → оценочная задача johndoe@gmail.com [Дом. страница асессора - недавно законченные задачи - выход]  
Язык: английский (US)  
**Оценочная задача - ICQ**  
[поисковые результаты: **google**]

Запрос	ICQ
Регион запроса	***Сан-Франциско***
Описание запроса	Это геозависимый запрос с регионом запроса, указанным выше
URL	http://www.mobicq.info/
Местоположение задачи	США
Язык задачи	Английский
Др. допустимые языки	Нет

## ГЛАВА II: ОЦЕНКА URL С УЧЕТОМ РЕГИОНА ЗАПРОСА

### 3. ПРИНЦИП ОЦЕНКИ ПРИ НАЛИЧИИ РЕГИОНА ЗАПРОСА

В некоторых случаях регион запроса играет важную роль при присвоении оценки. Например:

**Запрос:** [Теремок рестораны], русский (RUS)

**Регион запроса:** Нальчик, Кабардино-Балкария

Регион запроса важен. Пользователи из Нальчика, которые набрали запрос, ищут рестораны «Теремок» в районе Нальчика, а не в других областях.

Тем не менее во многих задачах запрос не ассоциируется с определённым местоположением, поэтому регион запроса не будет приниматься во внимание. Оценка будет такой же, как и в случае, если бы задача не имела региона запроса. Например:

**Запрос:** [wildberries.ru], английский (US)

**Местоположение запроса:** Ольгинка, Краснодарский край

Местоположение запроса абсолютно не важно. Wildberries.ru сайт, который никак не связан с определённым местоположением.

Регион запроса важен в тех случаях, когда целевая страница может быть более ценной для пользователей в одной локации, чем в другой.

#### 3.1 КОГДА ВАЖЕН РЕГИОН ЗАПРОСА?

Вот несколько примеров, которые показывают, когда регион запроса играет важную роль, а когда нет.

Запрос	Регион запроса	URL	Вероятное намерение пользователя	Важен ли регион?	Объяснение
[facebook], английский (US)	Бузулук, Оренбургская область	<a href="http://www.facebook.com/">http://www.facebook.com/</a>	Пользователь в городе Бузулук Оренбургской области хочет зайти на сайт www.facebook. com	Нет, т.к. Фейсбук - сайт, который не привязан к местности	Целевая страница одинаково полезна как посетителям в городе Бузулук, так и в других местах. Страница оценивается как <b>Витальная</b> (Appropriate Vital) для любого региона.

Запрос	Регион запроса	URL	Вероятное намерение пользователя	Важен ли регион?	Объяснение
[якитория], русский (RUS)	Москва, Московская область	<a href="http://www.yakitoriya.ru/">http://www.yakitoriya.ru/</a>	Пользователь в Москве ищет информацию о ресторане «Якитория» в Москве или хочет зайти на сайт «Якитории».	Нет, т.к. домашняя страница для объектного запроса всегда получает оценку <b>Витальная</b> (Appropriate Vital), даже если существует страница для данного геозависимого запроса.	Официальная домашняя страница должна оцениваться как <b>Витальная</b> (Appropriate Vital) для Москвы и любого другого региона.
[якитория], русский (RUS)	Москва, Московская область	<a href="http://www.yakitoriya.ru(locations/moskva-zapad-msk-zap">http://www.yakitoriya.ru/locations/moskva-zapad-msk-zap</a>	Пользователь в Москве ищет информацию о ресторане «Якитория» в Москве или хочет зайти на сайт «Якитории».	Да, т.к. московским пользователям нужны рестораны «Якитория» именно в Москве, а не где-то ещё.	Целевой страницей является официальная страница ресторана «Якитория» в Москве. В данном случае странице присваивается оценка <b>Витальная</b> (Appropriate Vital). В случае если регион запроса не указан, страница оценивается как <b>Витальная</b> (Other Vital) или <b>Относительно релевантная</b> .
[каток], русский (RUS)	Тутаев, Ярославская область	<a href="http://www.mayakkatok.tut.ru/">http://www.mayakkatok.tut.ru/</a>	Пользователь в Тутаеве ищет информацию о местных катках.	Да, т.к. пользователи в Тутаеве ищут катки именно в Тутаеве, а не в других местах.	Целевой страницей является официальная страница ледового катка "Маяк" - единственного катка в Тутаеве. Это доминирующая интерпретация для запроса данного региона. В других регионах эта страница будет оцениваться как <b>Нерелевантная/ Неполезная</b> или <b>Относительно релевантная</b> , если регион в запросе не обозначен.

Запрос	Регион запроса	URL	Вероятное намерение пользователя	Важен ли регион?	Объяснение
[погода], русский (RUS)	Обнинск, Калужская область	<a href="http://gismeteo.ru/RUS/Kaluzhskaya Oblast/Obninsk.html">http:// gismeteo.ru/ RUS/ Kaluzhskaya oblast/ Obninsk.html</a>	Пользователь в Обнинске ищет информацию о местной погоде.	Да, т.к. пользователей в Обнинске, скорее всего, интересует погода именно в Обнинске, а не в других городах.	Целевая страница содержит информацию о текущих погодных условиях в Обнинске. Оценивается как <b>Полезная</b> . Страницу нужно оценивать как <b>Нерелевантная/Неполезная</b> для других регионов или <b>Относительно релевантная</b> , если регион не обозначен.

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

### 1. ЧТО ТАКОЕ ПОИСКОВЫЙ СПАМ?

Поисковый спам — термин для веб-страниц, которые разработаны веб-мастерами для обмана поисковых систем и привлечения пользователей на свои сайты. В данном руководстве поисковый спам будет называться «спам», а веб-мастера, которые используют сомнительные методы, — «спамеры».

В этой главе вы узнаете, как распознать некоторые из обманных методов. Страницам, на которых вы встретите использование таких методов, вы будете присваивать метки **Спам**. При этом стоит отметить, что раздражающие, замусоренные и низкокачественные страницы со множеством всплывающих объявлений не обязательно являются спамом.

#### 1.1 ВЗАИМОСВЯЗЬ ОЦЕНОК И СПАМА

В главе 1 данного руководства вы узнали, что целевые страницы оцениваются в соответствии с их полезностью для человека, ищущего конкретный запрос. Невозможно оценить страницу без понимания запроса.

Метка **Спам** не зависит от связи запроса и целевой страницы. Страница должна получить эту метку, несмотря на свою полезность по данному запросу, (если она создана с использованием обманных методов).

Некоторые спам-страницы очень низкого качества и почти не содержат полезных сведений. В дополнение к метке **Спам** этим страницам присваиваются низкие оценки, такие как **Относительно Релевантная** и **Нерелевантная/Неполезная**.

Спам-страницам не самого низкого качества и с каким-либо полезным контентом может быть присвоена оценка **Относительно Релевантная** или **Релевантная**.

В некоторых особых случаях страница может получить оценку **Витальная** и при этом иметь метку **Спам**. Например, если есть скрытый редирект, но целевая страница является целью запроса — страница получает оценку **Витальная** и метку **Спам**. О скрытом редиректе вы узнаете из раздела 2.3 данной главы.

#### 1.2 ЗАЧЕМ СОЗДАЮТ СПАМ-СТРАНИЦЫ?

Спамеры обычно создают спам-страницы, чтобы заработать. Например, размещают платные ссылки на спам-странице.

Два основных типа платных ссылок (moneymaking links):

- **Pay-per-click-реклама** (PPC),
- **Аффилиаты** (представлены в разделе 4.1 данной главы)



Понравилась книга? Расскажи друзьям!

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

В случае с PPC-рекламой владелец сайта получает плату за каждый клик по объявлению. Другим названием PPC-рекламы является sponsored links. PPC-реклама встречается на очень многих веб-страницах. Одни из них — спам, другие — нет. Страницы не должны помечаться меткой **Спам**, если они содержат полезную информацию. Страницы являются спамом, если они бесполезны и созданы исключительно для заработка.

Иногда спам-страницы не содержат платных ссылок. Такие страницы создаются для влияния на выдачу поисковых систем либо для заражения пользовательских ПК вирусами. Это спам-страницы, т.к. они используют обманные методы.

### 1.3 КОГДА ПРОВЕРЯТЬ СПАМ

Страницы известных сайтов, например, <http://www.apple.com>, казалось бы, не нужно проверять на спам. Тем не менее даже веб-мастера авторитетных сайтов используют обманные методы. Для определения спамности страниц рекомендуем использовать два простых способа:

- При помощи нажатия «Ctrl-A» (или «⌘» и «A» для компьютеров Apple) проверьте целевую страницу на скрытый текст. Об использовании «Ctrl-A» вы узнаете из раздела 2.1.1 данной главы.
- Прокрутите страницу вниз и направо для просмотра скрытого текста вне поля зрения. Подробней об использовании скрытого текста вне обычного поля зрения вы узнаете из раздела 2.1.5 данной главы.

### 2. ПОИСК ТЕХНИЧЕСКИХ СИГНАЛОВ

При оценке страницы на спам нужно начинать с поиска следующих «технических сигналов»:

- Скрытый текст и скрытые ссылки.
- Перенасыщенность ключевыми словами.
- Скрытые редиректы.
- Клоакинг с использованием JavaScript-редиректов и 100% фреймов.

Данный раздел описывает эти технические сигналы и указывает способы их определения.

#### 2.1 СКРЫТЫЙ ТЕКСТ И СКРЫТЫЕ ССЫЛКИ

Веб-мастера добавляют скрытый текст и/или скрытые ссылки, чтобы завлечь поисковые системы и пользователей на свои страницы. Скрытый текст виден поисковикам, но не пользователям, которых он может отвлекать и раздражать. Вот что надо знать о скрытом тексте:

- Может быть не виден глазу.
- Может совпадать с цветом фона страницы или лишь немного отличаться от него.
- Может иметь самый мелкий формат (вплоть до 1 пикселя) и быть незаметным.
- Может находиться вне поля зрения. Например, на странице может быть большое пространство между обычным полем зрения и «скрытой» зоной текста в самом низу страницы или далеко справа.



Понравилась книга? Расскажи друзьям!

- Скрытый текст может состоять всего из пары строк или занимать целую страницу.
- Скрытый текст главным образом используют для обмана поисковых систем. Но порой он может не являться спамом. Например, если веб-мастер просто прячет дату обновления.

Как выявить скрытый текст:

- При помощи «Ctrl-A» (или «⌘» и «A» для Apple).
- Отключить CSS.
- Отключить JavaScript.
- Просмотреть код страницы.
- Проверить области страницы за пределами поля зрения.

## 2.1.1 ПРИМЕНЯЕМ «CTRL-A» НА ЦЕЛЕВОЙ СТРАНИЦЕ

После того как вы перешли на страницу, одновременно нажмите на клавиши «Ctrl» и «A» или «⌘» и «A» / «Command» и «A» (для Apple) и затем прокрутите вниз всю страницу. Данный способ позволяет выявить скрытый текст.

Мелкий текст не всегда удаётся распознать при помощи «Ctrl-A». Следует внимательно присматриваться к тексту страницы, потому что он иногда содержит скрытый текст. Простой способ распознать подвох — выделить и скопировать подозрительную строку, вставить в текстовый редактор и увеличить размер шрифта. Также можно использовать способы, приведенные ниже.

## 2.1.2 ОТКЛЮЧИТЬ CSS

Отключение CSS иногда позволяет распознать скрытый текст. Не обязательно таким образом проверять каждую страницу. Но если страница подозрительная — это надо сделать.

## 2.1.3 ОТКЛЮЧИТЬ JAVASCRIPT

Иногда спамеры используют JavaScript, чтобы разместить скрытый текст на странице.

**Важно:** Когда вы закончите проверять страницу, не забудьте включить JavaScript. Если этого не сделать, некоторый функционал целевых страниц не будет работать.

## 2.1.4 ПРОСМОТРЕТЬ КОД СТРАНИЦЫ

С помощью анализа кода страницы также можно выявить скрытый текст. Поиските в коде большие зоны, перенасыщенные ключевыми словами. Подробно про перенасыщение ключевыми словами будет рассказано в разделе 2.2 данной главы.

Обратите внимание, что метка **Спам** не присваивается, если перенасыщение ключевыми словами встречается только в метатегах. Метатеги легко распознаются, данная область начинается словами «meta name».

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

### 2.1.5 ПОИСК ЗА ПРЕДЕЛАМИ ОБЫЧНОГО ПОЛЯ ЗРЕНИЯ

Обращайте внимание на пустые области внизу и с правой стороны страницы. Пользуйтесь горизонтальной и вертикальной прокруткой для поиска скрытого текста за пределами поля зрения.

### 2.2 ПЕРЕНАСЫЩЕНИЕ КЛЮЧЕВЫМИ СЛОВАМИ

Порой веб-мастера наполняют страницу большим количеством ключевых слов. Вот как это выглядит:

- Ключевые слова часто повторяются на странице.
- Слова, идущие в связке с ключевыми, часто повторяются на странице.
- Множество неправильных написаний ключевых слов на странице.
- Страницы с большим количеством как бы случайных ключевых слов.
- Страницы, состоящие из сгенерированного бессмысленного набора слов.

Также веб-мастера наполняют страницу ключевыми словами и темами, не имеющими отношения к запросу, например, ипотека, сотовые телефоны, рингтоны, азартные игры, погода и т.д. Спамерам не важно, имеют ключевые слова отношение к теме или не имеют, цель — привлечь на страницу роботов поисковых систем и пользователей.

Порой трудно определить, является ли страница перенасыщенной ключевыми словами. Вы должны присвоить странице метку **Спам**, если думаете, что количество ключевых слов чрезмерно велико и будет раздражать/отвлекать пользователя. Если вы так не считаете — не присваивайте метку **Спам**.

Помните: скрытый текст и перенасыщение ключевыми словами обычно идут в связке. Скрытый текст зачастую содержит чрезмерное количество ключевых слов.

#### Определение перенасыщения ключевыми словами

В некоторых случаях перенасыщение текста ключевыми словами заметно на глаз, и не обязательно применять особые приёмы для его выявления. Иной раз текст, перенасыщенный ключевиками, скрыт. Обнаружить скрытое перенасыщение ключевыми словами можно с помощью метода, изложенного в разделе 2.1 данной главы.

**Важно:** скрытое перенасыщение ключевыми словами **всегда считается спамом (исключением являются метатеги).**

#### 2.2.1 ПЕРЕНАСЫЩЕНИЕ КЛЮЧЕВЫМИ СЛОВАМИ В URL

Ссылки также могут содержать чрезмерное количество ключевых слов. Такие ссылки генерированы автоматически на основе слов запроса и часто состоят из дефисов и пробелов. Это является явным сигналом, что ссылка спамная.

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

Вот несколько примеров перенасыщения ссылки ключевыми словами. Необязательно переходить по таким ссылкам, чтобы убедиться в спамности страницы.

- <http://frat-boy-gay.grandbrooklynlodge/boy-brief-frat-in-their-wet.html>
- <http://brazilian-model-alexandra.wantloweryour.cn/brazilian-model-adriana-lima.html>

### 2.3 СКРЫТЫЕ РЕДИРЕКТЫ

Мы говорим о скрытом редиректе в том случае, когда страница перенаправляет пользователя с URL на одном домене на URL на другом домене. Поисковые системы «видят» первую страницу, в то время как пользователь перенаправляется на другую страницу с другим контентом.

Вот что ещё следует знать о скрытых редиректах:

- Редирект проходит через несколько URL-ов, пока не приведёт на целевую страницу.
- Скрытые редиректы могут вести пользователя на разные меняющиеся домены; таким образом, кликнув на одну и ту же ссылку несколько раз, вы всякий раз попадаете на новую целевую страницу.
- Некоторые скрытые редиректы ведут на известные коммерческие веб-сайты, например, Amazon, eBay, Zappos и др.

#### Определение скрытых редиректов

- **Сравните две ссылки:** Сравните URL-ы страниц, участвующих в редиректе (откуда и куда перенаправляется пользователь). Перенаправление со старой страницы сайта на новую не является скрытым редиректом. Перенаправление с одной страницы на другую страницу одного и того же домена также не считается скрытым редиректом.
  - **Посмотрите, на кого зарегистрирован домен:** Если вы подозреваете скрытый редирект, проверьте сведения о владельце обоих доменов. Если владелец один, то вряд ли это скрытый редирект. Инструкции по определению владельца домена ищите в разделе 2.3.1 данной главы.

#### 2.3.1 ИСПОЛЬЗОВАНИЕ СЕРВИСА WHOIS

Вот как это делается:

##### 1. Используйте следующие сайты:

<http://www.domaintools.com/> и <http://whois.mtgsy.net/default.php>. Некоторые компьютеры позволяют ввести команду типа «whois example.com» через службу терминалов в Windows.

2. Вводите URL домена в окно поиска на странице Whois. Возможно, придётся удалить некоторые слова до или после доменного имени. Например, если URL <http://supportapj.dell.com/support/>, то в строку поиска надо вводить только «dell.com».

##### 3. Открываете еще одну страницу сервиса Whois.

##### 4. Вводите URL другого домена в поисковую строку.

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

**5. Сравниваете регистрационные данные двух URL-ов.** Если они зарегистрированы одним лицом, можно сделать вывод о том, что это не спам. А если разными и сходства не видно, значит, спам.

**Обратите внимание,** что домены одного владельца могут выглядеть по-разному. Например, книжный магазин «Москва» может иметь следующие домены: [www.moskvakniga.ru](http://www.moskvakniga.ru), [www.mkniga.ru](http://www.mkniga.ru) и [www.knigi.ru](http://www.knigi.ru).

### 2.4 КЛОАКИНГ

Процесс, когда веб-мастер показывает поисковой системе одну страницу, а пользователю другую, называется клоакингом. Сам по себе клоакинг используется редко, обычно спамеры прибегают к другим методам, чтобы показать разные страницы поисковикам и пользователям. Вот два таких метода:

- JavaScript-редиректы,
- 100% фрейм.

#### 2.4.1 JAVASCRIPT-РЕДИРЕКТЫ

Спамеры используют редиректы на JavaScript, чтобы показать поисковой системе одну страницу, а пользователю — другую. Чтобы выявить данный спам-метод, нужно загрузить страницу сначала с JavaScript, а потом без него. Если использованы JavaScript-редиректы, разница будет видна.

#### 2.4.2 100% ФРЕЙМ

Часто веб-мастера маскируют то, что предназначено для «пользовательского глаза», с помощью фреймов. Создаются два фрейма (две страницы), но при этом один из них занимает 100% экрана. Пользователь видит один фрейм (страницу), тогда как поисковая система — оба.

### 3. ПОЛЕЗНЫЕ СТРАНИЦЫ VS СПАМ-СТРАНИЦЫ

Поисковые системы стремятся показывать посетителю полезные страницы. В этом разделе мы научимся отличать спам-страницы с рекламой от полезных страниц с рекламой. А также обсудим:

- Страницы с PPC-рекламой и другим содержанием, которые помогают пользователю.
- Страницы с PPC-рекламой и другим содержанием, которые предназначены, прежде всего, для зарабатывания денег.

Некоторые страницы содержат только PPC-рекламу либо имеют очень мало полезной информации кроме неё. Такие страницы являются «чисто рекламными».

О них рассказывается в разделе 4.2 данной главы. Если страница с PPC-рекламой имеет некоторую важность для пользователей, это не спам. Вот примеры полезного контента:

- **Возможность сравнивать цены:** Некоторые страницы предлагают пользователям сравнить

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

цены перед покупкой. Даже если пользователь должен уйти на другую страницу для оформления заказа, функционал для сравнения цен на странице — полезный сервис.

• **Обзоры товаров:** Некоторые страницы предоставляют отзывы о товарах, помогающие пользователю в выборе. Обычно отзывы оставляют о книгах, электронике или гостиницах.

• **Рецепты:** Если рецепты оригинальные и содержат отзывы, то страница не спам.

• **Стихи, цитаты, пословицы, тексты песен и т.д.:** Если такие страницы помогают пользователям найти нужные стихи или тексты, то это не спам.

• **Контакты:** Если страницы сообщают контактную информацию: адреса, номера телефонов, схему проезда и т.д., то страница не является спамной.

Обратите внимание, что рецепты, стихи, цитаты, и т.д. обычно публикуются на не очень авторитетных страницах.

### 3.1 СТРАНИЦЫ С КОПИРОВАННЫМ ТЕКСТОМ И РРС-РЕКЛАМОЙ

Копированный контент — это содержание, скопированное с других источников. Веб-мастера зачастую пользуются программой-скрепером, которая выбирает в сети тексты с нужными ключевыми словами, и размещают найденные тексты на своей странице. Также контент можно просто скопировать с другого сайта.

#### 3.1.1 КОПИРОВАННЫЙ ТЕКСТ И РРС-РЕКЛАМА

Контент, взятый из открытых источников, таких как «Википедия» (<http://www.wikipedia.org/>) или каталог DMOZ (<http://www.dmoz.org/>), всё равно является копированным.

Копирование контента из таких источников не является plagiatом и не запрещено. Веб-мастера, которые копируют контент, обычно не упоминают его настоящих правообладателей и приписывают авторство себе. Даже если указаны авторы, текст является копированным.

Копированный контент обычно устаревший, необновляемый, в некоторых случаях ему нельзя доверять. Пользователи же ищут проверенную информацию. Копия статьи из «Википедии» на неизвестном сайте в окружении рекламы бесполезна для посетителя. Такую страницу следует считать спамом, т.к. она содержит копированный контент и создана для зарабатывания на размещённой рекламе.

#### 3.1.2 ФИДЫ И РРС-РЕКЛАМА

На сайтах (таких как BBC, РБК, Usenet, Cnet, РИА Новости, «Атом» и др.) публикуется информация, которая распространяется среди пользователей по каналам RSS и XML. Такие компании, как, например, Searchfeed.com, предоставляют фиды с РРС-рекламой и ссылками на опытных веб-мастеров. Страница, содержащая фиды и РРС-рекламу, созданная исключительно в коммерческих целях, является спамом.

#### 3.1.3 ДОРВЕИ

Дорвеи — это набор страниц, созданных для поисковых систем, который перенаправляет пользователя на одну общую целевую страницу. Большинство страниц выглядит одинаково и не даёт пользователю внятной информации. Например:



Понравилась книга? Расскажи друзьям!

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

<http://www.hair-removal-hair-laser.com/>. Эта страница содержит ссылки для всех Штатов Америки. Нажимая на ссылку, пользователь надеется получить страницу с информацией по конкретному региону, но, если нажать на другую ссылку, становится понятно, что все страницы одинаковые. Такие страницы являются спамом.

### 3.1.4 ШАБЛОНЫ И ДРУГИЕ АВТОМАТИЧЕСКИ СГЕНЕРИРОВАННЫЕ СТРАНИЦЫ

Некоторые сайты содержат шаблоны для массового создания страниц. Контент обычно копируется из доступных источников. Вы научитесь распознавать шаблоны, которые обычно повторяют исходный формат. Ищите ключевые слова в разных формах, не сильно измененных. Например, если ключевое слово «mortgage», то в заголовке, тексте и/или в доменном имени вам встретятся следующие его формы: «mortgages», «mortgage loan», «mortgages loans» и т.д.

Такие спам-страницы содержат ссылки на другие страницы с разным сочетанием копированного контента, PPC-рекламы и других спам-ссылок. Переходя по ссылкам, вы попадёте на другие страницы одного и того же домена с аналогичным контентом и ссылками.

### 3.1.5 КОПИРОВАННЫЕ ДОСКИ ОБЪЯВЛЕНИЙ

Иногда попадаются копированные доски объявлений. Если страница содержит только копированные доски объявлений и PPC-рекламу — это спам.

### 3.1.6 ОПРЕДЕЛЕНИЕ КОПИРОВАННОГО КОНТЕНТА

Данная информация поможет вам распознать копированный контент:

- **Поиск точного совпадения по целому предложению:** Скопируйте и вставьте определённое предложение в строку поиска. Закавычте данное предложение, чтобы найти точное совпадение. Можно вставить предложение без знаков препинания и других символов, чтобы не осложнять ими поиск полного совпадения. По найденным вариантам можно будет понять происхождение контента. Если контент оригинальный и ниоткуда не скопирован, скорее всего, он создан, чтобы помочь пользователю.
- **Ищите PPC-рекламу, которая окружает контент:** «Википедия» и DMOZ не показывают рекламу. Если на странице контент из «Википедии» или DMOZ в окружении PPC-рекламы — страница спамная.
- **Ознакомьтесь с форматами страниц в «Википедии» и DMOZ:** Заголовки разделов и ссылки на страницах «Википедии» обычно одного формата. Страницы DMOZ используют легко распознаваемую навигационную цепочку. Кроме того, на страницах DMOZ встречаются ссылки «подтвердить сайт» и «стать редактором», которые могут присутствовать на скопированных страницах.

### 3.1.7 ДРУГИЕ СПОСОБЫ ОПРЕДЕЛЕНИЯ СПАМ-СТРАНИЦЫ

Вот дополнительные способы определения спамных страниц:

- **Посмотрите на формат URL:** Формат URL подскажет, был ли использован шаблон или

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

другой автоматический способ для его создания. URL-ы такого типа часто содержат ключевые слова, перечисленные через дефис. Например: <http://nzeland.co.nz/blog/thelawmail/2007/12/29/com-search-extreme-belladonna-users-search-expireddomain-names-search-expired-domains/>.

- **Оцените, полезна ли страница посетителю:** Особое внимание обратите на страницы со стихами, рецептами, цитатами, контактной информацией, номерами телефонов, адресами, оригинальными отзывами, окном для комментариев и другими элементами, оживляющими текст.

- **Обратите внимание на непонятные и автоматически сгенерированные словоформы:** Посмотрите на текст страницы. Обычно сгенерированный текст представляет собой набор несочетаемых слов и тарабарщину. В тексте много ключевых слов. Автоматически созданные страницы часто не имеют ценности для пользователя. Это спам.

### 3.2 ФЕЙКОВЫЕ СТРАНИЦЫ ПОИСКА С РРС-РЕКЛАМОЙ

Фальшивая поисковая страница — это страница, которая выглядит как оригинальная страница с поисковыми результатами, например, Google. На странице присутствует строка поиска, но если ввести туда новый запрос, то вы попадете на другую страницу со ссылками. Если нажать на некоторые — появляется с набором ссылок PPC-рекламы, замаскированных под результаты поиска.

### 3.3 ФЕЙКОВЫЕ БЛОГИ С РРС-РЕКЛАМОЙ

Фейковые блоги содержат фальшивые записи, которые либо бессмысленны, либо скопированы с другого источника. Такие блоги часто перенасыщены ключевыми словами, о которых рассказано в разделе 2.2 данной главы. Страница создана таким образом, что на PPC-ссылки пользователь обязательно кликнет. PPC-ссылки могут быть расположены в тексте либо в других частях страницы. Зачастую функция комментирования на фейковых блогах не работает. Фейковые блоги — это спам.

**Заспамленные блоги:** Такие блоги отличаются от фейковых. Заспамленный блог — это работающий блог с настоящими статьями, но перенасыщенный постами/сообщениями с PPC-рекламой или/и порноссылками. Мы не хотим наказывать блог за то, что кто-то другой разместил на нём спам. Если вы видите, что блог нормальный, а засорил его кто-то посторонний, не присваивайте метку **Спам**.

### 3.4 ФЕЙКОВЫЕ ДОСКИ ОБЪЯВЛЕНИЙ С РРС-РЕКЛАМОЙ

Фейковая доска объявлений аналогична фейковому блогу. Она содержит «сообщения», которые таковыми не являются. Текст сообщений может быть бессмыслицей или содержать ссылки на PPC-рекламу. Фейковые доски на первый взгляд содержат комментарии, форму регистрации, но они либо нерабочие, либо перенаправляют пользователя обратно на ту же страницу. На настоящих досках можно найти ответы на вопросы, посты. На фальшивых ответов либо нет, либо они спамные.

**Копированные доски объявлений с РРС-рекламой:** Существуют доски, которые полностью скопированы. Если у вас возникло такое подозрение, проверьте это, скопировав в поиск часть текста. Копированные доски объявлений — спам.



Понравилась книга? Расскажи друзьям!

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

**Заспамленные доски объявлений:** Такие доски отличаются от фейковых. Заспамленные доски — это настоящие доски объявлений с реальными постами и ответами, но замусоренные PPC-рекламой и/или порноссылками. Если видно, что доска объявлений настоящая, но замусорена спамом, не присваивайте метку **Спам**.

### 3.5 КОПИРОВАННЫЙ КОНТЕНТ, КОТОРЫЙ НЕ ЯВЛЯЕТСЯ СПАМОМ

Иногда копированный контент не является спамом. Вот некоторые примеры: стихи, поэмы, пословицы, цитаты и т.п. Такое содержание неуникально и не имеет правообладателя.

Если оцениваемая страница содержит стихи, поэзию и т.п. слегальных сайтов, не присваивайте ей метку **Спам**. Если видно, что страница создана в основном для заработка, метка **Спам** присваивается.

## 4. КОММЕРЧЕСКИЕ НАМЕРЕНИЯ

В этом разделе мы поговорим о том, как спамеры зарабатывают и как выявить коммерческие намерения.

Большинство спам-страниц имеют коммерческую направленность. Они создаются с целью заработать, например, на получении комиссий при переходе по PPC-ссылкам.

Если страница создана только ради получения денег и бесполезна для посетителей, она считается спамной.

**Запомните:** Некоторые спам-страницы могут не иметь явного коммерческого намерения. Часто их создают для влияния на поисковую выдачу или для распространения вирусов.

### 4.1 НЕПОЛНЫЕ АФФИЛИАТЫ

Аффилиат зарабатывает на комиссиях. Такие сайты создаются для получения прибыли. Аффилиат часто содержит информацию с других «настоящих» коммерческих сайтов, например, Amazon, eBay, какой-нибудь гостиницы или агентства путешествий. Когда пользователь кликает, чтобы сделать заказ, его перенаправляют на реальный сайт продавца.

**Неполный аффилиат** — это сайт, который предлагает лишь немного дополнительной информации, помимо скопированной, и поэтому не несет особой ценности для пользователя по сравнению с другими ресурсами в сети. Например, аффилиат, который лишь дублирует информацию с сайта продавца, — это неполный аффилиат, созданный для заработка. Такие страницы являются спамом.

#### 4.1.1 ВЫЯВЛЕНИЕ НЕПОЛНЫХ АФФИЛИАТОВ

Чтобы определить, что страница является неполным аффилиатом, нужно сделать следующее:

- **Проверьте содержимое страницы на оригинальность.** Качество страницы-аффилиата зависит от того, сколько полезной, уникальной/дополнительной информации по сравнению с другими ресурсами содержится на данной странице. Если она имеет такой же текст или функционал,

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

как и другие подобные сайты, это наверняка спам.

- **Проверьте владельцев домена.** Если, кликнув, вы попадаете на другую страницу — проверьте данные о владельце доменов. Если владелец один и тот же, скорее всего, страница не является неполным аффилиатом. Следуйте инструкциям по использованию сервиса Whois в разделе 2.3.1 данной главы.

### 4.1.2 ОПРЕДЕЛЕНИЕ НАСТОЯЩИХ ПРОДАВЦОВ

Параметры, помогающие определить сайты настоящих продавцов:

- Ссылка «Просмотр корзины» не уводит с сайта.
- Корзина обновляется при добавлении товаров.
- Возврат товара производится на физический адрес.
- Калькулятор доставки работает.
- Есть ссылка «Хочу купить/Вишлифт» или ссылка «Отложить покупку».
- Есть способ отследить посылку.
- Есть рабочий форум для пользователей.
- Есть возможность регистрации.
- Есть настоящий/работающий розыгрыш призов.

#### Имейте в виду:

- Не все из вышеперечисленных параметров должны одновременно содержаться на нормальной странице продавца.
  - Yahoo!Store — это реальные продавцы.
  - Некоторые продавцы могут перенаправлять пользователя на другой сайт для завершения оплаты. При этом они не являются неполными аффилиатами.

Многие крупные онлайн-продавцы (например, Amazon.com, eBay.com, Zappos.com, AllPosters.com, Hotels.com, Orbitz.com и Overstock.com) предлагают аффилиат-программы.

### 4.2 СТРАНИЦЫ ТОЛЬКО С РРС-РЕКЛАМОЙ

На страницах только с РРС-рекламой содержится совсем небольшое количество контента, и/или они полностью состоят из ссылок на рекламные объявления. Спамер зарабатывает при каждом переходе по данным ссылкам. Такие страницы могут включать ссылки на другие спам-страницы, которые тоже содержат РРС-рекламу. Страницы только с РРС-рекламой — это спам.

### 4.3 ПРИПАРКОВАННЫЕ/ПРОСРОЧЕННЫЕ ДОМЕНЫ

Слово «домен» имеет два разных значения для асессора:

- Это может быть один из элементов системы доменных имен.
- А также это может быть набор слов (URL), который идентифицирует веб-адрес определённой компании, например, microsoft.com, harvard.edu, baidu.cn и т.п.



Понравилась книга? Расскажи друзьям!

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

В этом разделе слово «домен» используется во втором значении.

Когда компании выбывают из бизнеса, приобретаются другими организациями, меняют название, не платят за доменное имя, домен считается просроченным и может быть продан любому желающему.

**Припаркованные домены:** Спамеры покупают просроченные домены с истекающим сроком оплаты и помещают на их страницах свой контент. Их ценность для спамера в имеющихся ссылках, которые ведут на данный сайт с других ресурсов. Ведь теперь все эти ссылки ведут на страницу спамера.

Спамеры приобретают следующие типы доменов, которые мы также относим к припаркованным по причине их сходства:

- Домены, приближенные по написанию к оригинальным в надежде на то, что пользователь ошибется в написании имени домена и будет направлен на их веб-сайты, содержащие PPC-рекламу.
- Домены с именами, которые может набрать пользователь при поиске определенного сайта.

Типичный припаркованный/просроченный домен содержит все пункты или часть из них:

- Список платных ссылок.
- Список популярных категорий.
- Список категорий, который содержит ключевые слова.

### Выявление припаркованных/просроченных доменов:

• **Посмотрите на ссылки.** Обычно все ссылки на припаркованном домене платные. На странице практически отсутствует оригинальный контент.

• **Посмотрите на имя домена (URL).** На припаркованном домене имя (URL) практически никогда не связано с содержанием веб-страницы.

• **Посмотрите на страницу в интернет-архиве.** Зайдите на <http://www.archive.org/index.php>, введите URL и просмотрите, как выглядела страница, когда принадлежала первому владельцу. Если исходный сайт отличается, скорее всего, это припаркованный домен.

Уже скоро вы познакомитесь с припаркованными/просроченными доменами на практике.

## 4.4 СТРАНИЦЫ С БЕСПОЛЕЗНЫМ СОДЕРЖАНИЕМ И РРС-РЕКЛАМОЙ

Некоторые страницы создаются для того, чтобы размещать на них рекламу. Спамеры платят копирайтерам за написание текстов на различные темы. Зачастую статьи выглядят искусственно и не содержат полезной информации, но при этом они оригинальные. Таких нет на других сайтах. Однако явно видно намерение ввести пользователя в заблуждение. Но если контент оригиналный и со смыслом — метка **Спам** странице не ставится.

- Посмотрите на контент и задайте себе следующие вопросы:
  - Контент полезный или размытый, плохо написанный или бессмысленный?



Понравилась книга? Расскажи друзьям!

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

- Ценна ли страница на фоне других результатов поиска?
- Не слишком ли много на странице рекламы, которая отвлекает и сбивает пользователя?
- Вы бы поверили содержанию?
- Оставили бы вы реквизиты своей платёжной карты на этом сайте?

- Постарайтесь определить, кем создана страница — человеком или машиной.
- Постарайтесь определить, зачем создана страница.

### 5. ФИШИНГОВЫЕ САЙТЫ

Фишинг — это попытка бессовестных людей выудить конфиденциальную информацию у интернет-пользователей. Возможно, некоторые из вас получали письма, которые выглядят как официальные, но при ближайшем рассмотрении таковыми не оказываются. Часто в них содержатся просьбы поделиться конфиденциальной информацией.

Целевая страница в нижеприведенной задаче запрашивает конфиденциальную информацию и является еще одним типом фишинга.

**Запрос [runescape gold], English (US)**

**URL <http://www.grunescape.com/>**

Такая целевая страница должна сильно насторожить пользователя (и асессора). Орфография и грамматика неадекватны, страница похожа на спамную. Но самое ужасное — она может требовать ввести банковский пароль и пин-код!

Хотя мы вряд ли введём туда свои данные, такой тип фишинга не противоречит «Руководству по поисковому спаму» и такой странице нельзя присвоить метки **Спам** или **Вредоносная**.

Такими метками помечайте только те типы страниц, которые перечислены в «Руководстве». Фишинговые страницы могут быть спамом, но в данном случае это не они.

### 6. СПАМ И СТАДИЯ РЕШЕНИЯ

Нередко оценочные задачи попадают в стадию решения, потому что асессоры расходятся во мнении присвоить странице оценку **Нерейтингуемая: Не загрузилась** или другую оценку и метку **Спам**. Расхождения возникают, потому что асессоры видят разные страницы при переходе по ссылке. Это происходит из-за времени, версии браузера и/или настроек.

Когда задача попадает в стадию принятия решения по одной из этих причин, вам нужно повторно взглянуть на нее. И если другие асессоры воспринимают страницу как спамную, очевидно, что они видят что-то отличное от того, что видите вы. Тут можно попробовать следующее:

- Обновить браузер до последней версии.
- Посмотреть исходный код страницы или отключить JavaScript.

## ГЛАВА III: РУКОВОДСТВО ПО ПОИСКОВОМУ СПАМУ

Если спама нет — не присваивайте метку **Спам**.

Имейте в виду, спамные страницы часто не загружаются по прошествии определённого времени. Если вы нашли спам однажды, а на следующий день страница не загружается, не меняйте свою оценку (т.е. не убирайте метку **Спам**).

### 7. ЗАКЛЮЧЕНИЕ

Определение спама — это навык, который вырабатывается на практике в реальных условиях.

Не забывайте смотреть на страницу в целом. Спам-страницы обычно имеют некоторые из следующих характеристик:

- PPC-ссылки очень выделяются на странице, и явно заметно, что страница была создана под них.
- Если проверить контент страницы, вы увидите, что текст скопирован.
- Если условно закрыть все спам-элементы страницы (PPC-рекламу и копированное содержание), ничего ценного не останется.

Качественные страницы обычно обладают следующими характеристиками:

- Хорошо организованы. Реклама может быть, но она хорошо различима и не отвлекает.
- При проверке текста первой выдаётся данная/оригинальная страница.
- Страница полезна. Поисковая система отображает страницу в выдаче.

Метки **Спам**, которые вы будете использовать:

- **Не спам:** Если видите, что страница качественная, присваивайте метку **Не спам**.
- **Возможно, спам:** Если вы сомневаетесь в том, что данная страница спамная, присваивайте метку **Возможно, спам**.
- **Спам:** Если видно, что страница была создана с использованием сомнительных методов, описанных в данном руководстве, присваивайте метку **Спам**.

Когда не уверены, какую метку присвоить, задайте себе следующие вопросы:

- Нашёл ли пользователь, что искал на этой странице?
- Содержит ли страница оригинальный контент, ценный для пользователя?
- Будет ли включена такая страница в результаты поиска?
- Страница сделана для пользователя? Её создавал человек или машина?
- Если убрать PPC-рекламу и копированный текст, останется ли что-нибудь полезное?

Если да, то, скорее всего, страница не заслуживает метки **Спам**.